THE MANIPULATION OF USER EXPECTANCIES: EFFECTS ON RELIANCE, COMPLIANCE, AND TRUST USING AN AUTOMATED SYSTEM

A Thesis
Presented to
The Academic Faculty

by

Andrew K. Mayer

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Psychology

Georgia Institute of Technology
May, 2008

THE MANIPULATION OF USER EXPECTANCIES: EFFECTS ON RELIANCE, COMPLIANCE, AND TRUST USING AN AUTOMATED SYSTEM

Approved by:

Dr. Arthur D. Fisk, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Wendy A. Rogers
School of Psychology
Georgia Institute of Technology

Dr. Gregory M. Corso
School of Psychology
Georgia Institute of Technology

Date Approved: 03-25-08

For Kate

ACKNOWLEDGEMENTS

I would like to thank Drs. Fisk and Rogers for their support and guidance. I would also like to thank all the members of the Human Factors and Aging lab. This research was supported in part by a grant from the National Institutes of Health (National Institute on Aging) Grant P01 AG17211 under the auspices of the Center for Research and Education on Aging and Technology Enhancement (CREATE). This research was also supported in part by contributions from Deere & Company. I would like to thank Jerry Duncan and Bruce Newendorp for their support and advice on this research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

As automated technologies continue to advance, they will be perceived more as collaborative team members and less as simply helpful machines. Expectations of the likely performance of others play an important role in how their actual performance is judged (Stephan, 1985). Although user expectations have been expounded as important for human-automation interaction, this factor has not been systematically investigated. The purpose of the current study was to examine the effect older and younger adults' expectations of likely automation performance have on human-automation interaction. In addition, this study investigated the effect of different automation errors (false alarms and misses) on dependence, reliance, compliance, and trust in an automated system.

Participants interacted with an Automated Warehouse Management System (AWMS) during four experimental blocks. During the first three blocks the AWMS was 90 percent reliable and during the fourth block it was 60 percent reliable. The automation committed either false alarms or misses. Expectancies were manipulated by providing participants with a written description of the AWMS that framed likely performance of the automation in terms of high, low, or standard performance.

Overall, younger adults with high expectancies depended on, relied on, and complied with the automation more than participants with low and standard expectancies; however, the effect lasted only through the first block. Older adults did not exhibit significant expectancy effects. These results are important because expectancies may have to be considered when designing training for human-automation interaction.

During the first three blocks, participants in the false alarm condition increased reliance and decreased compliance whereas participants in the miss condition did not change behavior. When transferred to the less reliable version of the automation (i.e., Block 4), younger adults in the false alarm condition reduced both reliance and compliance whereas younger adults in the miss condition reduced only reliance. Understanding the effects of automation error type is crucial for the design of automated systems. For example, if the automation is designed for diverse and dynamic environments where automation performance may fluctuate, then a deeper understanding of automation functioning may be needed by users.

CHAPTER 1

INTRODUCTION

"Confident expectations of a certain intensity or quality of impression will often make us sensibly see or hear it in an object which really falls far short of it." (James, 1890/ 1981, p. 402)

Automation has made many of our everyday tasks safer and more efficient (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003). Banking can be conducted faster and at our convenience, GPS navigation systems can guide us to our destination, and collision-avoidance systems warn us if we are about to collide with a hazardous object. Automation has also had an impact in the most complex systems such as nuclear power plants and airplane cockpits (Wickens & Xu, 2002). The goals of introducing automation, to increase performance, safety, and efficiency on tasks, can be accomplished by reducing operator workload and subsequent cognitive demands (Dixon, Wickens, & McCarley, 2007).

Cognitive declines related to aging have been well documented (e.g., Craik & Salthouse, 2000). Reaction time, fluid abilities, visual search, divided attention, and working memory capacity have all been shown to change with age (e.g., Gottlob, 2006; Salthouse, 1992; Siedlecky, Salthouse, & Berish, 2005). These natural declines have been shown to have effects on performance of many tasks, including activities of daily living (ADLs), which can lead to a loss of independence for older adults (Clark, Czaja, & Weber, 1990). Given the goals of automation: to increase performance, safety, and efficiency on tasks; it seems that older adults could benefit greatly from the use of automated systems (Fisk & Rogers, 2002). It is therefore imperative that we study automation with both younger AND older adult user populations.

In both simple and complex systems, rarely is the operator/user or the automation subservient to the other, rather, they work as a "team" toward common goals (Bowers, Oser, Salas, & Cannon-Bowers, 1996). Like any team or group interaction, having an understanding of group capabilities and limitations can help gauge the potential performance of the group (Bowers et al.; Stephan, 1985). Understanding how the automated system, as a team member, is likely to perform, and the aspects of the task that it is likely to perform well, provides the operator with a basis for appropriate use (Cohen, Parasuraman, & Freeman, 1998). Appropriate use involves depending on the automation at times when it is providing correct information and disregarding the automation when it is not providing correct information. Therefore, for successful human-automation interaction the user must appropriately depend on the automation.

1.1 Dependence, Reliance and Compliance

Traditionally, researchers studying human-automation interaction have considered human usage of the automation in terms of *dependence* on the system. Dependence is a global measure of automation usage across given states of the automation (Dixon & Wickens, 2006). Recent investigations, however, have dissected user dependence into two distinct constructs: reliance on and compliance with automation (Dixon & Wickens; Meyer, 2001; 2004). Reliance occurs when the automation is silent, not asking for an action from the human. In the case of an automated warning device, reliance refers to the expected behavior of the user during the *non-alarm state*. Compliance refers to performing the action asked for by the automation. Again, considering an automated warning device, compliance refers to the expected behavior of the user during the *alarm state* of the automation.

To illustrate the distinction between compliance and reliance, imagine driving down the highway in your car equipped with a collision avoidance system. You begin to change lanes, without quickly looking over your shoulder or glancing in your mirrors. The system suddenly warns you that someone is in your blind spot. If you heed the warning and return to your lane you have complied with the system. In contrast, if you begin to change lanes and the system provides no warning and you complete the lane change, again without shoulder checking or glancing in your mirrors, you have relied on the system.

The type of automation error interacts with compliance and reliance behaviors. Behavior can be different depending on whether an imperfect automated system commits errors of incorrect rejection (misses) or errors of incorrect rejection (false alarms, e.g., Dixon & Wickens; Dixon, Wickens, & McCarley, 2007; Meyer).

1.2 Automation Error Type

The type of error an automated system is likely to make depends on the response criterion. A more conservative criterion will result in the system providing fewer false alarms at the expense of committing more misses (Wickens & Carswell, 2006). A miss occurs when the system fails to detect the signal in the environment. In contrast, a more risky criterion will result in the system committing fewer miss events at the expense of committing more false alarms. A false alarm is the false detection of a signal that is not present in the environment (Wickens & Carswell). Misses occur when the system is inappropriately not providing any warning whereas false alarms occur when the system is inappropriately providing a warning (Meyer, 2001). Because these error types are distinct and influence distinct states of interaction, appropriate interaction with a system depends on the criterion of the system and users' understanding of that criterion.

Younger and older adults can and do adjust their behavior depending on the automation's criterion (Sanchez, 2006). Participants who interacted with a system that only committed errors of incorrect detection (false alarms) changed their behavior such that they almost always relied on the automation but almost never complied with the automation. In contrast, participants who interacted with a system that committed errors of incorrect rejection (misses) almost never relied on the automation but almost always complied with the automation (Sanchez). Age-related differences were found in terms of the length of time it took to adjust behavior and the degree of that adjustment. Overall, older adults were slower to adjust their behavior and when the adjustment did occur, it was to a lesser degree than younger adults.

Type of automation errors and age-related differences in dependence and performance has been investigated. Johnson (2004) tested younger and older adults using a dual-task, flight simulation where participants had to monitor engine gauges and a radar screen. The engine gauge task was aided by an 80% reliable automated system that, depending on condition, provided mostly false alarms, mostly misses, or an equal mix of both error types. Overall, older adults depended on the automation less than younger adults. In addition, within each age group, there were no statistically significant differences in dependence between the three failure conditions, although numerically, participants in the majority miss condition relied on the automation more than participants in the other two failure conditions.

Overall, younger adults outperformed older adults on the task, which was measured by combining the score on the radar task and engine monitoring task (Johnson, 2004). Within each age group, participants in the FA condition had the lowest overall performance, suggesting that false alarms may be more detrimental than misses on performance.

Misses and false alarms have been shown to have differential effects on reliance and compliance behavior (Meyer, 2001). Meyer showed that an increase in misses only reduces reliance whereas an increase in false alarms only reduces compliance. However, more recent evidence suggests that reliance and compliance may not be the independent constructs as originally proposed by Meyer (Dixon, Wickens, & McCarley, 2007). Dixon, Wickens, and McCarley found that a system with increased misses reduced reliance, but seemed to have no effect on compliance. However, they found that a system with increased false alarms reduced compliance, but also reduced reliance suggesting that false alarms have qualitatively different effects on dependence than misses. Finally, they found that false alarms had quantitatively different effects on performance, by reducing reliance and compliance, compared to misses (Dixon, Wickens, & McCarley).

In the study by Sanchez (2006), participants were provided with examples of potential system errors that informed participants of the difference between a false alarm and a miss. The instructions provided were neutral to reliability in the sense that participants were not told what type of error would be committed by the system. This type of instruction is standard in many automation studies and, although they are neutral to reliability, they do provide users with a rudimentary understanding of system states and functioning. However, during interactions with unfamiliar automated systems, people may not have access to this type of instruction and subsequent understanding.

There is some evidence that participants fail to appropriately calibrate their behavior according to the response criterion of the automation when a complete understanding of automation errors (false alarms and misses) is unavailable (Mayer, Sanchez, Fisk, & Rogers, 2006). Appropriate use of automation is aided by knowledge and understanding of the automation's capabilities and limitations but a deep level of understanding may be unavailable during interactions with unfamiliar automated systems. In such cases, expectations of automation capabilities and limitations act as a form of understanding that will likely influence how users will interact with automation.

1.3 User Expectancies

An important component of appropriate dependence involves processing information that provides the user with an understanding of the automations' capabilities and limitations (Cohen et al., 1998; Lee & See, 2004). There are a number of ways a user can develop such an understanding. For example, experience with the automation, which can guide future interactions with the automation, will provide the user with information regarding its reliability and performance (Chappell, 1997; Riley, 1996; Sanchez, 2006). However, with the proliferation of automated systems, individuals often face situations requiring interaction with unfamiliar systems. In such cases, users may only have expectations of likely performance to guide dependence. Such expectations, acquired through product marketing, product documentation, related experiences, or from other users, may guide dependence behavior.

Expectancies bias information selection and often lead people to focus attention on information that confirms held expectations (e.g., Jamieson, Lydon, Stewart, & Zanna, 1987; Snyder & Frankel, 1976). In the context of automation, expectations may guide operators to select expectancy-confirming information that may guide operators' use of the system. Thus, expectancies may critically affect dependence on automated systems.

The influence of expectancies on behavior has been extensively investigated in the social cognition literature (Stephan, 1985). Expectancies bias people by priming certain information and, often unconsciously they seek out and more fully process information that is consistent with their expectancies, further strengthening those expectations (e.g., Rosenthal, 1966; Rosenthal & Jacobson, 1968). This effect of expectancies was illustrated in a study conducted by Cantor and Mischel (1977). Participants were told that they would be learning about either extroversion or introversion. Participants then read ten statements about either extroverted or introverted people. After reading the statements, participants' were presented with traits that were included or not included among the statements. Participants who expected to be learning about extraverts were more confident that they had seen non-included descriptors of extraverts compared to participants who expected to be learning about introverts (Cantor & Mischel).

Rothbart, Evans, and Fulero (1979) found they could bias participants' judgments about a group's "intelligence" by telling the participant they would be interacting with a group of people who were either intelligent or friendly. If expecting interaction with an intelligent group, participants rated the group more intelligent compared to participants who expected a friendly group of people. These two studies are not isolated in terms of their findings (e.g., Feldman & Theiss, 1982; Harris & Rosenthal, 1985; Jamieson, Lydon, Stewart, & Zanna, 1987).

Little research has evaluated expectancy as a moderator of optimal automation usage. Expectancies, related to how an automation aide will perform, translate to a priori predictions of when and how well the automation will accomplish its role. Expectations can develop from prior experiences with the system or a similar system. Expectations also may develop from generally held biases or through information acquired from other users, advertisements, user manuals, or any number of other sources. The two studies reviewed next did investigate the role of expectations on human-automation interaction.

Effects of user expectations on human-automation interactions. Using an airport security luggage screening task, Madhavan and Wiegmann (2007) investigated the role of pedigree (whether the automation was labeled as an ‘expert’ or a ‘novice’) on reliance and compliance. There was no difference between the ‘novice’ and ‘expert’ automation other than the description of the automation that was provided to participants. During the task, participants had to indicate whether a hidden weapon (e.g., a knife) was present or absent in various x-ray luggage images. The automated aid provided participants with a determination of whether a hidden weapon was present or absent.

Madhavan and Wiegmann (2007) found that pedigree had no effect on reliance and compliance when the automated aid was 90 percent reliable, but when the aid was 70 percent reliable, participants initially relied on and complied with the expert automated aid more than the novice aid. However, after interacting with the imperfectly reliable system, participants who had the expert aid reduced their reliance and compliance.

The authors proposed that the reduction in reliance and compliance was due to a violation of a 'perfect automation schema' suggesting that development of trust in automation is different than development of trust in humans (Madhavan & Wiegmann, 2007). Participants had an expectation of a particular level of performance that when violated by system errors caused a reduction in reliance and compliance. A limitation of this study was that the system committed both false alarms and misses so the differential effects of error type could not be assessed.

In a dual-task scenario, younger adult participants were provided with a high or low expectancy of likely automation performance (Mayer et al., 2006). When interacting with a 90% reliable system, participants in the high expectation group depended on the automation significantly more often than participants in the low expectation group. In fact, participants in the high expectation group over-depended on the automation (depending about 98% of the time) whereas participants in the low expectation group under-depended on the automation (depending about 74% of the time).

Although few studies have specifically investigated the role of user expectancies, some studies have shown indirect evidence of the importance of expectations on human-automation interaction. Using a failure detection task, Chappell (1997) examined automation usage across "experienced" and "inexperienced" participants. Participants gained experience by interacting with a 100% reliable version of the system one day prior to the critical testing day. The inexperienced group interacted with the automation for the first time on the critical testing day. On the critical test day all participants interacted with a 90% reliable system. Fewer experienced participants were able to detect the first failures compared to inexperienced participants which may be due to expectations developed during the experience-acquisition day. The previous day's experience may have carried over into the critical test day reducing those participants' ability to detect the initial automation failures.

Wickens, Helleberg, and Xu (2002) illustrated the influence of pilot expectations on reliance behavior. The purpose of their study was to investigate pilot maneuvering tendencies when faced with conflicting air traffic types. In that study they found that pilots who were told that the automated system was imperfect showed smaller reductions in reliance following system errors compared to pilots who were not told of the automation's fallibility. Trust in the automation may guide people's dependence behavior when no prior understanding of the automation available. The next section explores the construct of trust as it relates to human-automation interaction.

1.4 Trust in Automation

Trust in a particular system's automation is a subjective measure of a user's confidence in that automation (Wiegmann, Rich, & Zhang, 2001). A person's trust in the automation accrues as that automation behaves in both a predictable manner and in the best interest of the user (i.e., that the action of the automation will be beneficial to the goals of the operator) (Muir, 1994). Trust in the automation does not always correspond to the users' dependence on the system (Lee & Moray, 1994; Wiegmann, Rich, & Zhang, 2001), however, trust in automation will guide users' dependence when a complete understanding of the task being automated is too complex or unavailable to the user (Lee & See, 2004).

A number of characteristics and factors that contribute to trust have been identified, including the reliability of the automation within the system. Reliability of automation, in the most general sense, can be defined as the total number of automation responses less the number of automation errors divided by the total number of automation responses. However, moving beyond this simple definition of reliability is required to understand how automation errors influence trust and subsequent reliance on the automation because type of error (miss or false alarm) has a significant affect on trust and reliance (Dixon & Wickens, 2004; Johnson, 2004; Sanchez, 2006; Wickens & Dixon, 2002).

The effects of system reliability and age differences in trust and subjective reliability have been investigated (Sanchez, Fisk, & Rogers, 2004). Older adults were found to be more sensitive to changes in system reliability compared to younger adults. Specifically, younger adults perceived no difference between a 60% and an 80% reliable system whereas older adults did. The same effect was found for subjective measure of trust. In addition, differences in reliance between the 80% and 60% reliable systems were greater for older adults compared to younger adults, although this effect was not significant (Sanchez et al.). In the study by Johnson (2004) described previously, older adults in the miss condition had significantly greater trust in the system compared to older and younger adults in all other conditions.

Trust, when used in the context of automation can be a challenging hypothetical construct. A dictionary definition of trust is “(a) assured reliance on the character, ability, strength, or truth of someone or something; or (b) one in which confidence is placed; dependence on something future or contingent; a charge or duty imposed in faith or confidence or as a condition of some relationship” (Merriam-Webster, 2006). Such definitions of trust are vague, but they all imply an expectation held by the person who has ‘trust.’ In fact, definitions of trust used in the social sciences emphasize expectation as a primary component of trust (Barber, 1983). Concepts such as predictability, reliability, and dependability have been identified as important factors of trust (Larzelere & Houston, 1980).

An important component of trust is the expectation held by the person trusting (Barber, 1983). Barber stated “In my exploration of the meanings of trust I start with the expectations that actors have of one another, because expectations can be thought of as the basic stuff or ingredient of social interaction, as matter is the basic stuff of the physical world” (p. 9). A necessary condition of trust appears to involve developing positive expectations of behavior for a group, person, or entity (Barber). Expectations influence trust as well as the information people seek and process when interacting in groups. Previous research on expectancy and research on automation discussed previously, suggested a qualitative model where expectancy exerts an influence on both trust and the selective information processing during interaction with the automation (Figure 1). In that model, trust in the automation and selective information processing due to expectancy combine to influence user dependence on the automation.

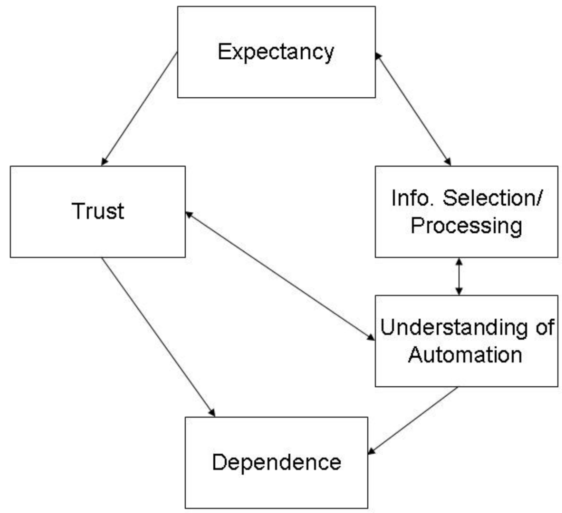


Figure 1: Basic conceptual model of relationship between expectancy and dependence. Causality, differential value of links, and so on should not be inferred from this figure.

1.5 Overview of the Study

The purpose of this study was to determine if user expectancies influence dependence on an automated system. In addition, reliance, compliance, and trust were investigated as a function of user expectancy and type of automation errors. Age-related differences were also investigated.

The social psychology literature indicates that expectancies bias people during interactions such that they focus attention on information that confirms their expectancies (Stephen, 1985). It is therefore expected that participants with high expectations will focus on instances the automation provides correct information leading to high dependence on the automation. In contrast, participants with low expectations will focus on instances the automation errs, leading to lower reliance on the automation. Participants in a standard group will adjust their reliance behavior more closely to the reliability of the automated system. Following from this, if the reliability of the system changes (i.e., if it becomes less reliable), the time required for participants to adjust behavior may be indicative of the information that is guiding participants' dependence behavior.

If participant expectancies are guiding behavior, we may expect participants with high expectancies to be slower to adjust their dependence behavior when the system shifts from being 90% to 60% reliable compared to participants who receive standard instructions. We would expect that participants with low expectancies would not adjust their overall dependence behavior because they would already be depending at about the same level as the reliability during the transfer block. However, we would expect the variability of dependence would be reduced as participants in the low expectancy group would have more opportunities to access expectancy confirming evidence. If participants are guiding their behavior by matching their dependence to the actual reliability of the system then we would expect participants with high and low expectancies to adjust their dependence behavior similarly to participants who receive standard instructions.

An important automation characteristic affecting reliance and compliance is type of automation error. Mixed findings have been reported in the literature so it is unknown how error type will affect reliance (e.g., Dixon & Wickens, 2004; Johnson, 2004; Sanchez, 2006). However, the salience of false alarms may make false alarm error information more available to participants (Dixon & Wickens) that may result in more extreme responses (Eisen & McArthur, 1979). With this in mind, it may be easier for participants in the low expectation group to selectively process error information (due to the salience of FA's) leading to lower reliance compared to participants with low expectancies interacting with a system that misses. It may be harder for participants in the high expectation group to selectively process non-error information leading to lower reliance compared to the high expectation group in the miss condition.

Studies show that users calibrate their reliance and compliance behavior over time depending on the response criterion of the system, when users have a basic understanding of system errors (e.g., Dixon & Wickens, 2006; Sanchez, 2006). However, preliminary data show that without standard instructions, participants do not appropriately calibrate reliance and compliance behavior. Rather they depend on the system in a non-systematic fashion. In the proposed study it is expected that participants who receive standard instructions will appropriately calibrate their behavior according to the system's response criterion as has been previously reported. In contrast it is expected that participants in the high and low expectancy groups will not successfully calibrate their behavior due to a lack of understanding related to system capabilities and limitations. If calibration does occur, it will occur later in the experimental session compared to the standard instruction group.

Trust in automation has been discussed as an important indicator of dependence on automation when the user lacks an understanding of how the automation works. The social psychological literature indicates that expectations are important in forming trust (Barber, 1983). It is difficult to predict how trust will differ between expectancy groups. Following from the social psychological literature it may be expected that participants in the high expectancy group would have greater post experiment trust in the automation compared to the low expectancy group.

Age-related changes have not been extensively studied in the automation literature. Results from Johnson (2004) suggest that older adults relied on automation more compared to younger adults when the system performs at 80% reliability. However, Sanchez et al. (2006) reported no significant age-related differences in reliance. Therefore, it is unknown if older adults depend more or less compared to younger adults. It is unknown how expectations will affect reliance, compliance, and trust for older adults, but given that older adults may be more sensitive to changes in automation reliability (Sanchez et al., 2004), it is expected that older adults will be more sensitive to inconsistencies between framed expectancies and the actual reliability of the system. Thus, high and low expectancy older adults will more accurately calibrate their reliance and compliance behavior following automation errors compared to younger adults. It is expected that older adults will more quickly adjust behavior during the transfer block compared to younger adults.

CHAPTER 2

METHOD

2.1 Participants

Sixty older adults (M = 70.73, SD = 3.59) recruited from the Atlanta area, and sixty younger adults (M = 20.53, SD = 1.69) recruited from the Georgia Tech community participated in this study. Older adult participants were compensated thirty dollars and younger adults were compensated two and half credits to be applied to undergraduate psychology classes for their participation in the study. Participants were tested for near and far visual acuity and were required to have a minimum corrected vision of 20/40. One younger adult participant was excluded due to visual acuity below 20/40.

2.2 Simulated Scenario

A dual-task Automated Warehouse Management System (AWMS) was developed and used for the experiment. The AWMS was programmed in JAVA Script and was displayed on a 17 inch monitor connected to 3.20 GHz Pentium 4 computers.

In the scenario, participants played the role of the shipping/receiving manager at a hypothetical warehouse. Participants were responsible for receiving shipments from incoming trucks and ensuring that departing trucks were dispatched with full loads. The goal of each task was to earn as many points as possible. Both tasks are described below followed by the point breakdown for each task.

Receiving packages task. The general idea for this task was that the warehouse manager (played by participants) had a list of packages that were expected for delivery at the warehouse (the receiving tag list). The packages on the receiving tag list were represented by barcode-like patterns. A package with a barcode-like pattern (the shipment tag) would be delivered to the warehouse and participants had to crosscheck the shipment tag with their receiving tag list by matching the barcode-like pattern of the shipment tag with the barcode-like patterns on the receiving tag list. The barcode-like patterns were developed from a random string of five ASCII symbols

Participants matched the shipment tag to the corresponding receiving tag in the list by using the up and down arrow keys on the computer keyboard to scroll through the receiving tag list. Participants then pressed the 'receive' key to select the matching receiving tag (the 'receive key was a keyboard key labeled with the word 'receive' in capital letters) (See Figure 2 for an example of the receiving packages task). If the correct receiving tag was selected, feedback indicating a correct response was provided, and a new shipment tag and a new receiving tag list were displayed. If the incorrect receiving tag was selected, feedback indicating an incorrect response was provided, and a new shipment tag and a new receiving tag list were displayed. If participants took longer than the allotted time to respond, feedback indicating a time-out was provided and a new shipment tag and receiving tag list were displayed. Older adult participants were given ten and a half seconds to make a response while younger adult participants were given seven seconds to respond. These response times were determined based on usability testing prior to commencing this study.

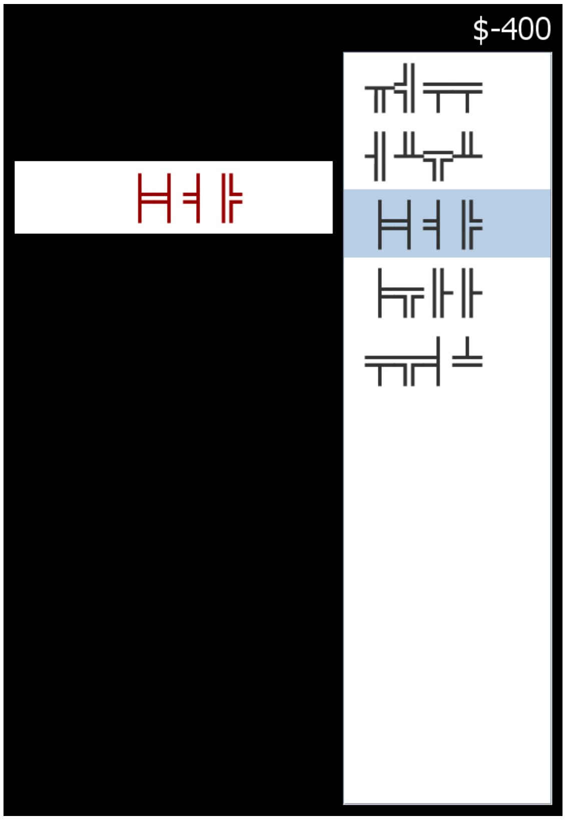


Figure 2. Example of receiving task. Participants were required to highlight the matching barcode-like pattern.

Truck-dispatching task. The dispatching trucks task required participants to dispatch fully loaded trucks from the loading dock. The dispatching trucks task was supported by an automated aid (the AWMS) that, conceptually, scanned the interior of the truck trailer and estimated when the truck was full. When the AWMS estimated that the truck was full, a visual notification was provided to the participant instructing them to dispatch the truck. Participants had to then press the ‘dispatch’ key (the ‘dispatch’ key was a keyboard key labeled with the word ‘dispatch’ printed in capital letters) to notify the truck driver to leave the loading dock. Participants had the option to check the automation by manually viewing the loading truck by pressing and holding the space bar key labeled ‘view truck’ (See Figure 3 for an example of the dispatching trucks task).

However, manually viewing the truck came with a cost; that is, the receiving packages task was hidden and the arrow keys required for the receiving packages task were disconnect while the 'view truck' key was being pressed.

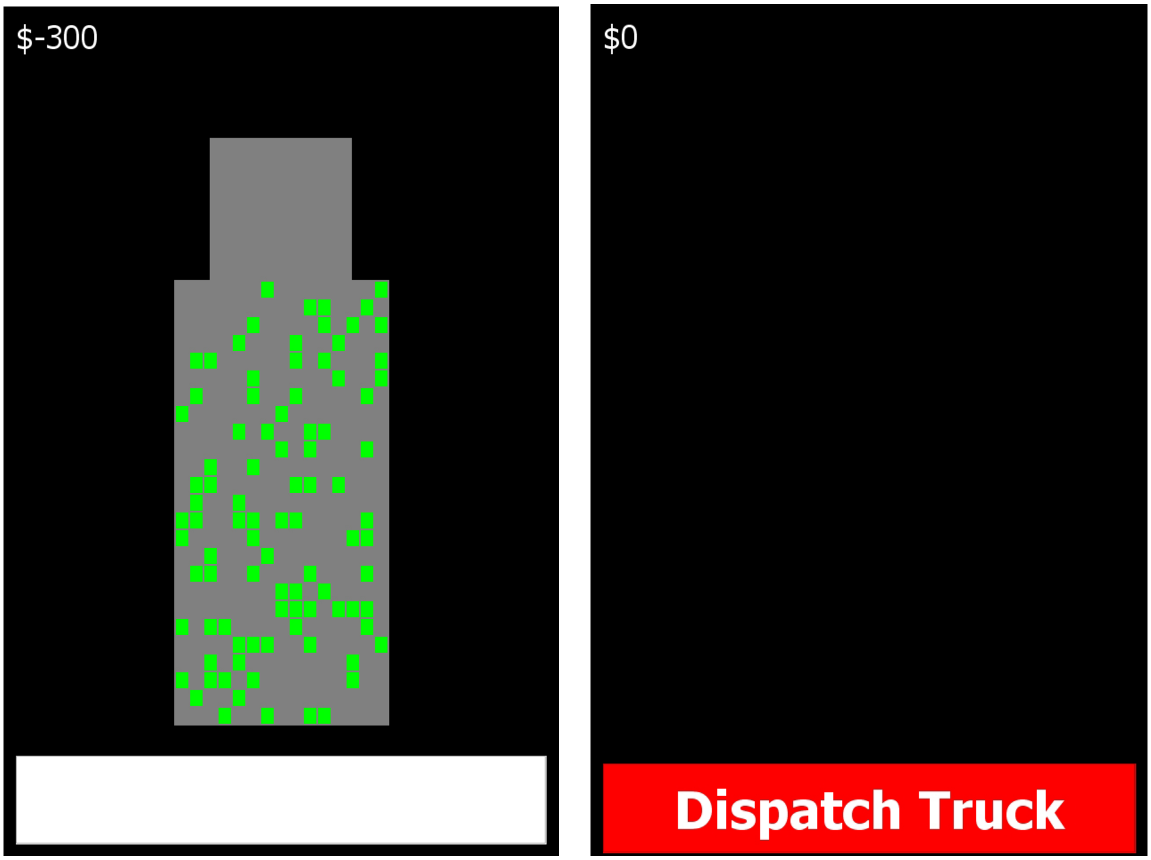


Figure 3. Left: example of truck filling as would be seen if the participant were pressing and holding the 'view truck' button. Right: example of the AWMS providing an automated alert indicating a full truck.

Trucks loaded at a random rate to prevent participants from estimating when a truck was fully loaded. Each truck took between 12 and 22 seconds to fill. Participants had 10 seconds to dispatch the truck once the truck was full. If the truck was not dispatched within 10 seconds of being filled, the truck overloaded and participants were penalized. If the truck was dispatched before it was full, participants were also penalized.

All participants managed 160 truck loadings, divided into four 40 truck blocks, such that the total experimental session lasted approximately 80 minutes.

Participants interacted with a system that either committed false alarms (FA) or misses. In the FA condition, the AWMS periodically indicated that the truck was fully loaded when in fact it was not. In the miss condition, the AWMS periodically failed to indicate that the truck was full when in fact the truck was full. In both the FA and miss conditions, the AWMS performed at 90 percent reliability during the first three experimental blocks and at 60 percent during the fourth block, the transfer block. Participants were not informed of what type of error the system would commit nor were they told the reliability of the system.

2.3 Point Scheme

Participants earned points for correctly receiving shipments and for dispatching fully loaded trucks. Participants lost points for incorrectly receiving shipments and for overloading trucks or sending incompletely filled trucks. Point totals were logged for the shipping and receiving tasks separately. In addition, a grand total was maintained for the combination of the shipping and receiving task points. The point tally for each task was shown throughout the experimental session. The point total for each task and the grand point total were presented at the end of each block. It should be noted that task performance, for the purpose of statistical analysis, was not measured in terms of points, but rather in terms of correct responses.

Receiving packages task. Because older adults generally required more time to make a response on the receiving packages task, they had to earn more points per correct response to equate their point performance with those of younger adults. This point scheme was determined during usability testing. Older adult participants earned 15 points for correct responses whereas younger adult participants earned 10 points for correct responses. Older adult participants lost 15 points for every incorrect response while younger adults lost 10 points for every incorrect response. If participants exceeded the time limit, the system timed out and was considered an incorrect response. The faster participants performed the receiving packages task, the more shipments that could be received and the more points that could be earned.

Dispatching trucks task. Participants received 100 points for dispatching a full truck. If participants dispatched a truck that was not full, they lost 200 points. If participants overloaded a truck, they also lost 200 points. The reason points were differentially rewarded and penalized was because initial usability testing with younger and older adults indicated that truck dispatching errors were largely ignored due to their infrequency and low cost. Increasing the cost to 200 points for incorrectly dispatched trucks was sufficient to make participants focus on both tasks equally.

2.4 Materials

Demographics. Demographic and general health information was gathered using the demographics questionnaire.

Ability tests. Participants completed the digit symbol substitution as a measure of general perceptual speed (Wechsler, 1997), the reverse digit span as a measure of general memory span (Wechsler), and the Shipley vocabulary test as a measure of general verbal ability (Shipley, 1986). The purpose of administering these abilities tests was to determine if groups differed within age and across conditions.

General trust and specific trust questionnaires. Prior to testing, participants completed a general trust questionnaire designed to evaluate general trust in automation. The general trust questionnaire was developed from questionnaires used by Sanchez (2006), and Jian, Bisantz, and Drury (2000) (See Appendix A for the general and specific trust questionnaires). The specific trust questionnaire was administered after participants completed the third experimental block. The specific trust questionnaire was designed to evaluate participant trust in the automated system with which they had just interacted.

Expectancy descriptions. Expectancy was operationally defined as the participant's belief regarding the automated system's likely performance during the experiment. Expectancy was manipulated by providing participants with a written description of the AWMS and the company that produced the system with which they would be interacting. The experimenter read the system description to participants while they followed along with a written version of the description. The description framed participants' expectations such that participants either expected good automation performance (high expectancy group) or poor automation performance (low expectancy group). A third expectancy group, the standard group, was told that the automated system was very reliable but could make errors. In addition, the standard group description provided participants with basic information regarding system misses and false alarms (See Appendix B for all three expectancy descriptions).

The high and low expectancy descriptions were manipulated along four categories: company history, type of technology, amount of system testing, and expected system performance. All other information provided in the descriptions was held constant for each manipulation. Table 1 provides a side-by-side comparison of the manipulation for the high and low expectancy groups for each category of manipulation. The standard group was designed to replicate instructions similar to those provided in previous automation studies (e.g., Sanchez et al., 2006) and was not manipulated along the same four categories as the high and low expectancy groups. In addition, the standard group provides data for comparison to findings from previous automation research.

A calibration study was conducted to test the effectiveness of the expectancy manipulations for the high and low expectancy conditions. A summary of the findings can be seen in Appendix C. Based on findings from the calibration study, it was expected that participants in the high expectancy condition would report higher expected system performance relative to the low expectancy condition.

Table 1

Comparison of High and Low Expectancy Descriptions

Category of Manipulation	**Low Expectancy (SRT-1) In-Text Manipulation**	**High Expectancy (SRT-2) In-Text Manipulation**
Technology	The company's first prototype system, the SRT-1, utilizes advanced decision algorithms and sensing technologies that have the ability to adjust to differing warehouse and loading conditions.	The company's latest groundbreaking system, the SRT-2, utilizes advanced decision algorithms and sensing technologies that have the ability to adjust to differing warehouse and loading conditions.
Company History	The company first became involved in sensory technologies in 2000 with the sole mission of creating advanced scanning and decision making systems for warehouse loading and shipping applications. In 2001, the company proposed an Automated Warehouse Management System called the SRT and in 2004 proposed a Smart Automated Warehouse Management System, the SRT -1.	The company first became involved in sensory technologies in 1975 with the sole mission of creating advanced scanning and decision making systems for warehouse loading and shipping applications. In 1985, the company released an Automated Warehouse Management System called the SRT and in 1997, released a Smart Automated Warehouse Management System, the SRT-1.
Level of System Testing	Testing of the SRT-1 has not begun so designers are unsure of how the accuracy, reliability, and robustness of the Automated Warehouse Shipping system will compare to the industry standard.	Testing of the SRT-2 indicates that it sets the industry standard for accuracy, reliability, and robustness and is still considered the leader in Automated Warehouse Management System systems.
Expected Performance	Because this is a first prototype Automated Warehouse Management System, it is expected that the SRT-1 will perform at a low level with some performance errors.	Because this is a well proven Automated Warehouse Management System, it is expected that the SRT-2 will perform at a high level with no performance errors

Expectancy questionnaire. Participants were provided with the respective expectancy manipulation description followed by an expectancy questionnaire designed to evaluate predictions of likely automation performance in the upcoming task. The Expectancy questionnaire can be seen in Appendix D.

2.5 Study Design

The level of user expectancy (low, high, or standard) and type of automation error (false alarm or miss) was manipulated as between participant variables. Exposure to the automation task was divided into blocks of time (~20 minutes per block) and was analyzed as a within participant variable. Age acted as a grouping variable. Each subgroup of age, younger and older adults was separated into six groups. Each group received the low, high, or standard expectancy manipulation and interacted with an automated system that either provided only false alarms or only misses. The reliability of the automated system was 90% during the first three experimental blocks. All participants were presented with a transfer block where the automation performed at 60% reliability.

2.6 Procedure

Upon receiving informed consent, participants completed the demographic and health questionnaire and the visual acuity tests.

Participants were then given a general definition of automation. Specifically, the definition read, “An automated system is a technologically-based system used to partially or fully assist the human in tasks involving sensing, detecting, information processing, making decisions and/or executing actions.” (Sanchez, 2005). After reading the definition, participants completed the general-trust in automation questionnaire.

Following the general-trust questionnaire, participants were provided with general information about large scale warehouse operations to provide context for the study (See Appendix E for Warehouse Operation description). Participants were then provided with their respective expectancy manipulation description. After receiving the description, participants completed the expectancy questionnaire. Participants had the written expectancy description in front of them while they completed the questionnaire.

Participants were then provided with detailed instructions of the experimental task. Following the instructions, participants completed four distinct practice blocks with the simulation. The first practice block allowed participants to practice only the receiving packages task without the time limit. Younger adults had to achieve 200 points and older adults had to achieve 300 points to move on to the second practice block, meaning, at a minimum, participants had to correctly receive 20 packages. The second practice block again allowed participants to practice only the receiving packages task but with the time limit restriction. Again, younger adults had to earn 200 points and older adults had to earn 300 points to move on to the next practice block. The third practice block allowed participants to practice only dispatching trucks. Participants had the opportunity to dispatch two trucks without the aid of the AWMS. The final practice block allowed participants to practice both receiving packages and dispatching trucks simultaneously. In the final practice block participants were aided by the AWMS (that performed at 100% reliability) and consisted of five truck-loading events for younger adults and seven truck-loading events for older adults.

The experimental session consisted of four blocks, each approximately 20 minutes in duration, separated by three breaks. Each break lasted a minimum of one minute but could last longer if participants needed more time. Participants completed the specific-trust questionnaire during the third break. When the specific-trust questionnaire was completed, participants completed the final block (the transfer block). In the transfer block, the AWMS committed the same errors (FAs or misses) as it did during the experimental session but only performed at 60% reliability. The purpose of the transfer block was to determine if participants were relying on expectancies to guide behavior or were matching their behavior to the reliability of the system. Participants did not receive any new instructions before beginning block 4.

After the transfer block, participants completed the reverse digit span, the digit symbol substitution, and the Shipley test. Finally, participants were debriefed and compensated for their participation.

CHAPTER 3

RESULTS

All tests were conducted using an alpha of .05. Follow-up tests were conducted using Fisher's LSD when comparing groups of three or corrected using a Bonferonni correction when comparing groups of more than three. The adjusted alpha level is reported for Bonferonni corrected tests.

3.1 Ability tests

Three ability tests, the Shipley Institute of Living Scale, the Reverse Digit Span, and the Digit Symbol Substitution, were administered to ensure participants fell within a normal expected range and to ensure equality between groups within age. Younger adults in each between subjects condition were statistically equivalent on all abilities. Older adults were equivalent on the reverse digit span and digit symbol substitution, but were slightly different on the Shipley vocabulary test. Older adults in the High Expectancy group had significantly higher scores on the Shipley compared to older adults in the Low Expectancy group ($t(9) = 3.15$). Overall, younger adults outperformed older adults on both the Reverse Digit Span ($F(1, 108) = 14.60$), and the Digit Symbol Substitution ($F(1, 108) = 66.45$). On the Shipley Institute of Living Scale, older adults scored better than younger adults ($F(1, 108) = 21.60$). A summary of the scores on the three ability tests for the 12 between subjects groups is presented in Table 2 & 3.

Table 2
Means and Standard Deviations for the Six Younger Adult Between Subjects Ability Scores

Condition	**Reverse Digit Span**		**Digit Symbol Sub.**		**Shipley**	
	Mean	SD	Mean	SD	Mean	SD
Low-FA	10.44	2.19	89.89	13.49	30.78	3.19
Low-Miss	11.20	1.75	82.80	14.88	32.80	3.12
High-FA	10.00	1.94	79.78	10.49	29.89	4.28
High-Miss	11.40	2.32	86.90	17.44	31.70	2.71
Standard-FA	9.75	1.91	87.00	9.38	30.75	5.15
Standard-Miss	9.10	5.04	87.20	9.91	31.70	2.06

Table 3
Means and Standard Deviations for the Six Older Adult Between Subjects Ability Scores

Condition	**Reverse Digit Span**		**Digit Symbol Sub.**		**Shipley**	
	Mean	SD	Mean	SD	Mean	SD
Low-FA	7.70	3.09	65.70	17.46	33.20	4.71
Low-Miss	7.80	2.90	53.60	10.23	32.80	4.13
High-FA	9.30	2.16	69.90	17.92	36.50	3.66
High-Miss	9.20	2.15	72.80	20.24	35.80	2.10
Standard-FA	7.70	2.41	50.50	15.00	35.10	4.25
Standard-Miss	8.80	2.25	67.40	13.21	33.10	3.25

3.2 User Expectancies

Recall, expectancies were operationally defined as participants' belief regarding the likely performance of the automation during the experiment and were manipulated by providing participants with a description of the automation. Following the description, the expectancy questionnaire was administered to determine the effectiveness of the manipulation.

The initial questionnaire, which assessed the immediate effect of the automation description on expected automation performance, demonstrated that the Expectancy manipulation was successful. On average, and as illustrated in Table 4, younger adults in the Low Expectancy condition expected the AWMS to provide correct information 70.2% of the time and older adults in the Low condition expected correct information 66.9% of the time. In contrast, in the High Expectancy condition, younger adults expected the system to be correct 93 % and older adults expected the system to be correct 90.6% of the time. Younger and older adults in the Standard condition reported expectancies of 89.3% and 81.9% respectively.

Table 4
Mean percent expectancy of participants by age and expectancy condition

Expectancy	**Error Type**	**Younger Adults**		**Older Adults**	
		Mean	SD	Mean	SD
Low					
	False Alarm	65.00	23.70	69.30	9.50
	Miss	75.30	11.40	64.50	25.10
High					
	False Alarm	93.30	7.70	87.40	15.20
	Miss	92.70	5.40	93.80	8.90
Standard					
	False Alarm	87.5	7.9	82.9	13.2
	Miss	91.1	8.1	80.9	29.7

There was a significant main effect of Expectancy ($F(2,114) = 23.74$, $p < .05$, $\eta^2 = .301$). This main effect was the result of lower expected automation performance of the participants in the Low Expectancy group compared to the other groups. Participants in the High Expectancy condition believed the automation accuracy to be higher than participants in the Low Expectancy condition, $t(9) = 23.27$, $p < .05$. In addition, participants in the Standard Expectancy condition believed the system would be more accurate compared with participants in the Low Expectancy condition, $t(9) = 17.08$, $p < .05$. There was no significant difference between the High and Standard Expectancy conditions, $t(9) = 6.19$, $p > .05$. Age did not produce different levels of expected automation performance, $F(1, 114) = 2.32$, $p > .05$, and Age did not interact with Expectancy, $F(2, 108) = .29$, $p > .05$.

3.3 Dependence, Reliance, and Compliance

The current study was designed to address specific questions related to human dependence, reliance, and compliance behavior. Recall that dependence is a global measure of automation usage, reliance is assessed when the automation is silent (non-alarm state), and compliance is assessed when the automation specifies an action (alarm state). Specifically, the goals in this regard were to understand how, across and within different age groups, user expectancies influence dependence and how expectancies interact with the type of system error to influence reliance and compliance behavior. The following analyses address these specific questions.

Dependence, as a global measure, does not differentiate between reliance and compliance and was measured by the participants' automation usage independent of the automation alarm state. Reliance was measured by calculating the percentage of time participants did *not* check whether the truck was full during the non-alarm state. Compliance was measured by calculating the percentage of time participants did dispatch the truck when the alarm-state was activated.

Previous research has suggested that when automation misses a state of the world that should signal an action then reliance on automation is reduced. Misses generally do not reduce compliance. Conversely, when automation falsely detects a state of the world demanding an action (a false alarm) generally compliance is reduced. Automation false-alarms either do not reduce reliance or reduce reliance but to a lesser extent compared with automation misses.

Expectancy and dependence on the automation during the average of blocks 1-3. Previous data suggest that user expectancies have a significant effect on dependence that was sustained for a 30 minute experimental session (Mayer et al., 2006). One goal for the current study was to investigate whether expectancy effects extended beyond this time. As such, the first three experimental blocks summed to a total of approximately 60 minutes and were first analyzed at the highest level, averaging across the three blocks.

The data are presented in Figure 4. Overall, older adults depended on the automation 73.9 percent of the time whereas younger adults only depended on the automation 45.5 percent of the time ($F(1, 114) = 33.94$, $p < .05$, $\eta^2 = .229$). There was a main effect of Expectancy ($F(2, 114) = 4.77$, $p < .05$, $\eta^2 = .077$). Participants with High Expectancies depended at a higher level than participants in the Standard Expectancy condition ($t(19) = 18.42$, $p < .05$). Numerically participants in the high expectancy condition depended on the automation at a higher level than participants in the low expectancy condition (68.5% and 60.6% respectively) but this difference was not statistically significant ($t(19) = 7.86$, $p > .05$). There was also no significant difference between the low and standard expectancy conditions ($t(19) = 10.55$, $p > .05$). The age by expectancy interaction was not significant ($F(2, 114) = .70$, $p > .05$).

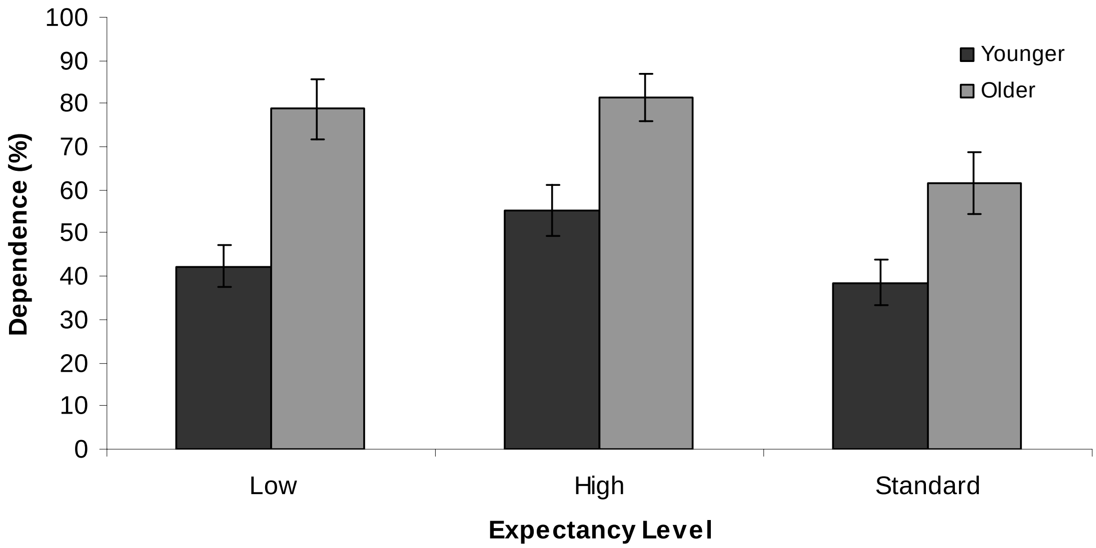


Figure 4. Dependence behavior by Age and Expectancy level averaged across the first three blocks.

Expectancy and dependence on the automation: A block-by-block analysis. Figures 5 and 6 illustrate each expectancy condition for younger and older adults during each of the first three blocks. As can be seen, for younger and older adults, High Expectancy led to the highest levels of dependence compared with the other Expectancy condition. Statistically, Expectancy effects were significant only for younger adults in block 1 ($F(2, 57) = 4.82$, $p < .05$, $\eta^2 = .145$). Younger adults in the High Expectancy condition depended on the automation 55.4 percent of the time compared to younger adults in the Low Expectancy and Standard Expectancy conditions who depended on the automation 40.5 and 38.6 percent of the time respectively ($t(9) = 17.50$, $p < .05$; $t(9) = 16.83$, $p < .05$). By Block 2, the effect was no longer statistically significant for younger adults ($F(2, 57) = 1.95$, $p > .05$, $\eta^2 = .064$), and by Block 3 the effect was even smaller ($F(2, 57) = 1.39$, $p > .05$, $\eta^2 = .046$). There were no significant expectancy effects for older adults in any of the three blocks (Block 1: $F(2, 57) = 2.89$, $p > .05$; Block 2: $F(2, 57) = 2.71$, $p > .05$; Block 3: $F(2, 57) = 2.29$, $p > .05$).

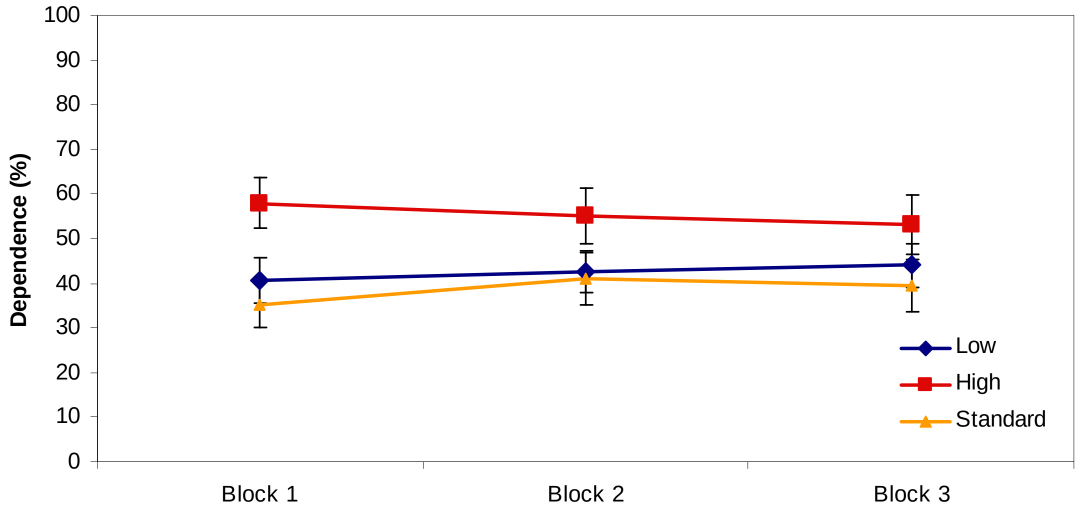


Figure 5. Dependence by Younger Adults in the Low, High, and Standard Expectancy conditions across Blocks 1, 2, and 3.

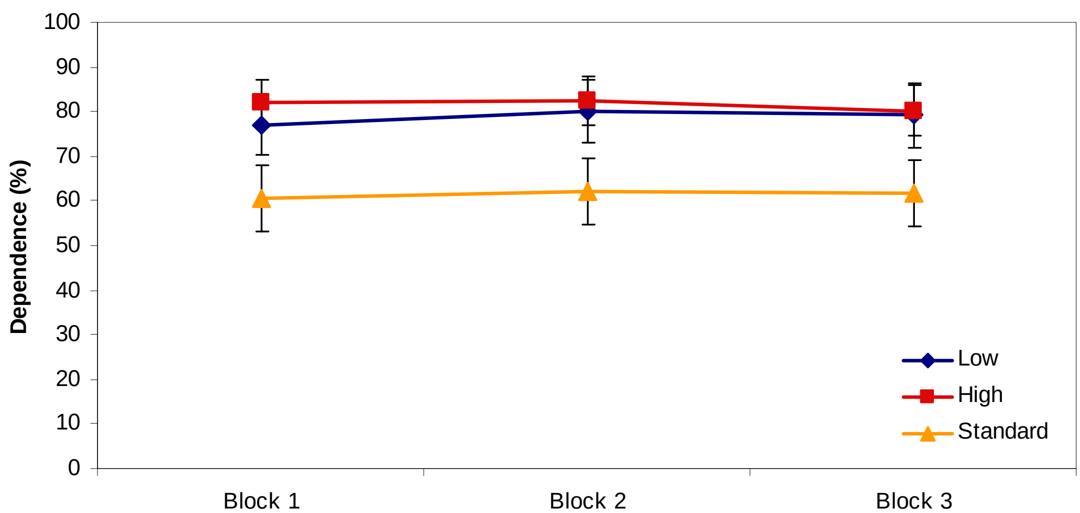


Figure 6. Dependence by Older Adults in the Low, High, and Standard Expectancy conditions across Blocks 1, 2, and 3.

Expectancy and dependence transferring from Block 3 (90%) to Block 4 (60%). Comparisons between Block 3 and Block 4 were used for the analysis of transfer.

Table 5 summarizes the means and standard deviations for Block 3 and Block 4 of each Expectancy group for both age groups. Figures 7 and 8 illustrate these data.

Table 5
Average Dependence by Younger and Older Adults in Block 3 and Block 4

Age	**Condition**	**Block 3 (90%)**		**Block 4 (60%)**	
		Mean (%)	SD	Mean (%)	SD
Younger Adults					
	Low	44.06	22.11	39.75	14.71
	High	53.13	29.95	45.13	21.75
	Standard	39.50	26.36	35.06	19.97
Older Adults					
	Low	79.13	32.2	76.88	33.84
	High	80.19	24.98	77.50	24.81
	Standard	61.81	33.57	61.44	31.37

As can be seen in Table 5, older adults depended on the automation to a greater degree than younger adults. It is clear that younger adults, in all three expectancy conditions decreased their dependence from Block 3 to Block 4 ($F(1, 57) = 7.75$, $p < .05$, $\eta^2 = .120$), and there was no significant expectancy effect for younger adults ($F(2, 57) = 1.52$, $p > .05$), or significant Block by Expectancy interaction ($F(2, 57) = .36$, $p > .05$). Older adults showed no decrease in their dependence from Block 3 to Block4 ($F(1, 57) = 2.56$, $p > .05$), and no effect of Expectancy ($F(2, 57) = 2.09$, $p > .05$), and the Block by Expectancy interaction was non significant ($F(2, 117) = .56$, $p > .05$).

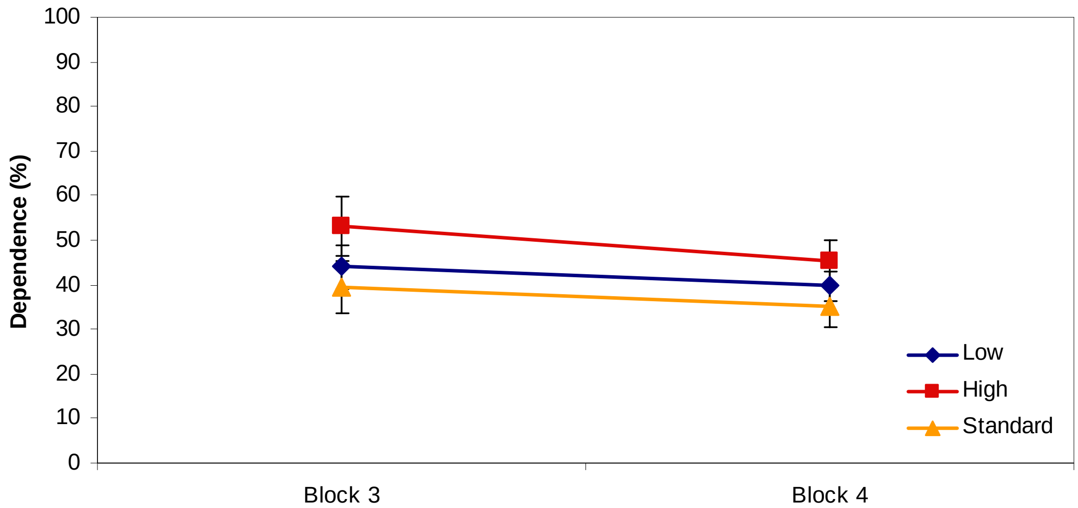


Figure 7. Dependence by Younger Adults in the Low, High, and Standard Expectancy conditions across Blocks 3 and 4.

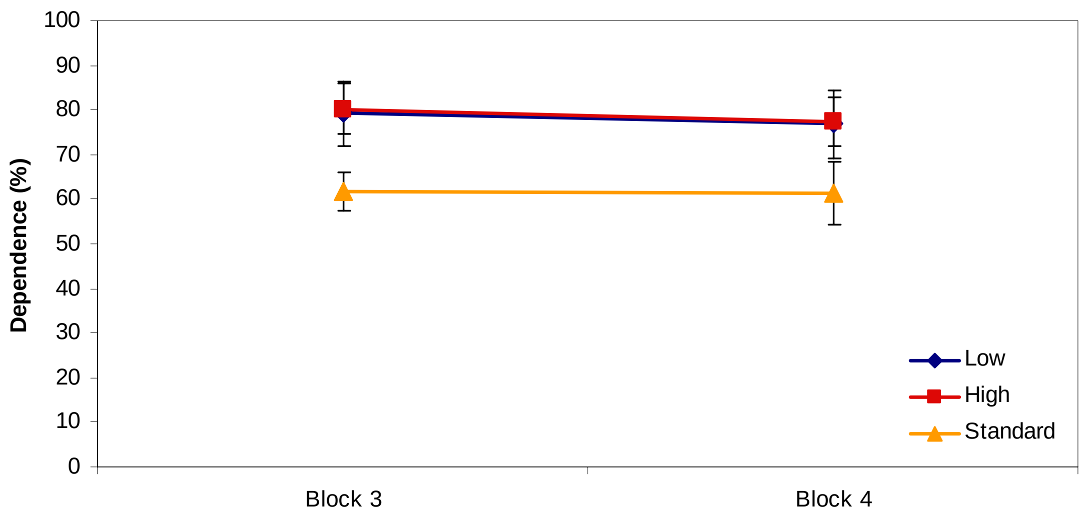


Figure 8. Dependence by Older Adults in the Low, High, and Standard Expectancy conditions across Blocks 3 and 4.

Expectancy, error type, and reliance on the automation during blocks 1, 2, & 3. Reliance behavior is presented in Figure 9 and represents checking on the automation when the automation was in the non-alarm state. Average reliance across the three blocks was 51.4, 57.7, and 58.0 percent respectively. This level of reliance is very low considering the system was 90 percent reliable during the first three blocks. The overall low reliance may be due to two possible factors: The pacing of the task and ease of checking the automation. Younger adults in particular relied at a very low level (35.8%) compared to older adults (75.8%) ($F(1, 108) = 44.62$, $p < .05$, $\eta^2 = .292$). Collapsed across Age, Error Type, and Block, the Standard Expectancy condition relied only 43.75 percent of the time compared to the Low, 56.77 percent, and the High Expectancy condition, 66.60 percent ($F(2, 108) = 4.95$, $p < .05$, $\eta^2 = .084$). The difference in reliance between the High and Standard conditions was significant ($t(9) = 22.85$, $p < .05$). There was a significant Block by Error Type interaction, $F(2, 107) = 5.91$, $p < .05$, $\eta^2 = .099$. Participants in the False Alarm condition did not rely heavily on the automation in Block 1 (46.2%) but significantly increased their reliance by Block 3 (60.7%) ($t(19) = 3.73$, $p < .0056$). In contrast, participants in the Miss condition maintained relatively stable reliance across the three blocks (54.6, 55.6, and 55.4 percent respectively).

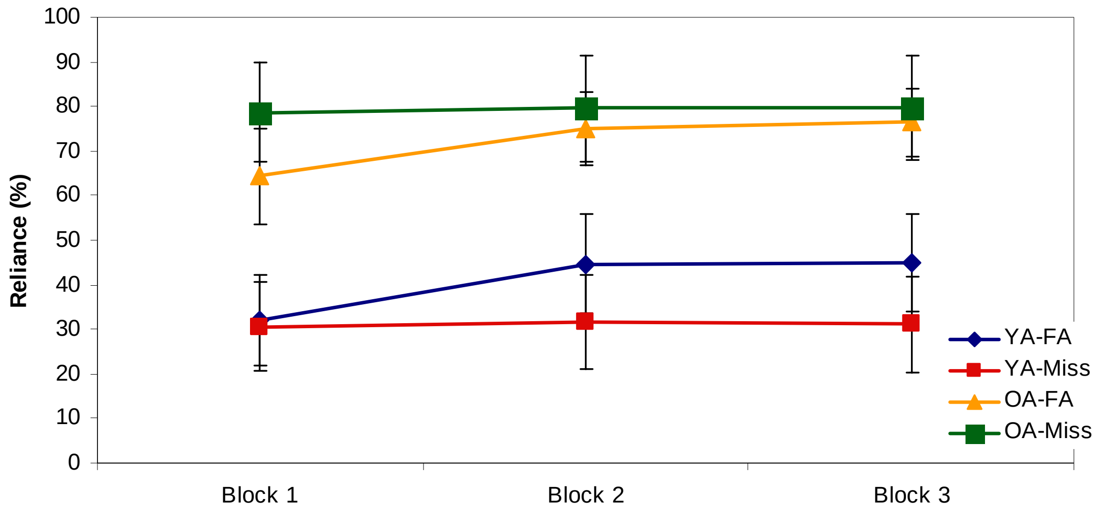


Figure 9. Reliance by Younger and Older Adults in the False Alarm and Miss conditions across Blocks 1, 2, and 3.

Expectancy, error type, and reliance on the automation: A block by block analysis. There was a statistically significant effect of expectancy for younger adults in Block 1 ($F(2, 54) = 3.20$, $p < .05$, $\eta^2 = .106$), but again, there was no significant effect of expectancy for older adults ($F(2, 54) = 2.29$, $p > .05$) (Figure 10). During Block 1, younger adults in the High Expectancy condition relied on the automation more than participants in the Standard Expectancy condition ($t(19) = 24.50$, $p < .05$). There was no significant effect of Error Type (miss or false alarm) during any block for younger or older adults.

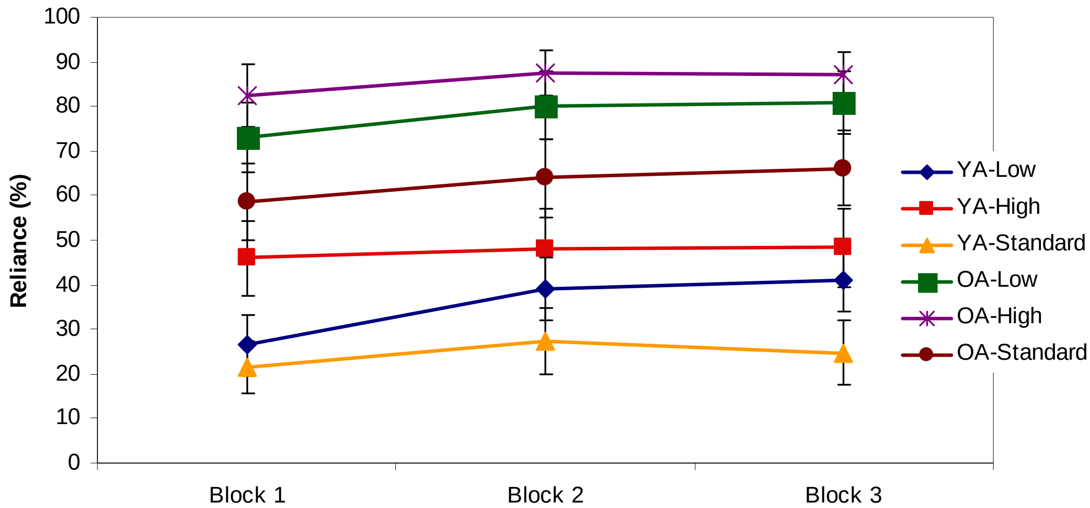


Figure 10. Reliance by Younger and Older Adults in the Low, High, and Standard Expectancy conditions across Blocks 1, 2, and 3.

Effect of transfer to Block 4 (60%) on reliance. Younger adults decreased their reliance from Block 3 to Block 4, regardless of Error Type or Expectancy as shown in Table 6 and Figure 11 ($F(1, 54) = 26.67$, $p < .05$, $\eta^2 = .331$). Overall, older adults also decreased their reliance from block 3 to transfer ($F(1, 54) = 20.00$, $p < .05$, $\eta^2 = .270$). However, a block by error type interaction was statistically significant ($F(1, 54) = 4.06$, $p < .05$, $\eta^2 = .070$) for older adults where older adults in the false alarm condition decreased their reliance but older adults in the miss condition did not. There was no such interaction for younger adults. The findings for younger adults are consistent with the findings from Dixon, Wickens, and McCarley (2007). The older adult data do not fit the Dixon et al. findings where it would be expected that both participants in the miss and false alarm conditions would decrease reliance. Yet, the older adult data do not fit the Meyer (2001) findings either, where it would be expected that only participants in the miss condition would decrease reliance. In fact, for the older adult data, only participants in the false alarm condition significantly decreased their reliance from Block 3 to Block 4.

Table 6

Average Reliance by Younger and Older Adults during Block 3 and Block 4

		Block 3		**Block 4**	
		Mean (%)	SD	Mean (%)	SD
Younger Adults					
False Alarm					
	Low	56.25	33.46	38.75	26.25
	High	55.75	37.97	28.50	23.61
	Standard	23.00	32.95	17.75	17.85
Miss					
	Low	26.00	29.37	13.75	12.37
	High	40.75	41.32	29.00	35.57
	Standard	26.50	30.85	18.25	29.46
Older Adults					
False Alarm					
	Low	72.00	35.08	66.00	39.21
	High	94.25	5.53	80.00	19.76
	Standard	63.00	31.55	48.25	27.49
Miss					
	Low	89.75	26.83	87.00	25.95
	High	80.00	39.72	73.25	37.27
	Standard	69.25	43.24	65.50	40.90

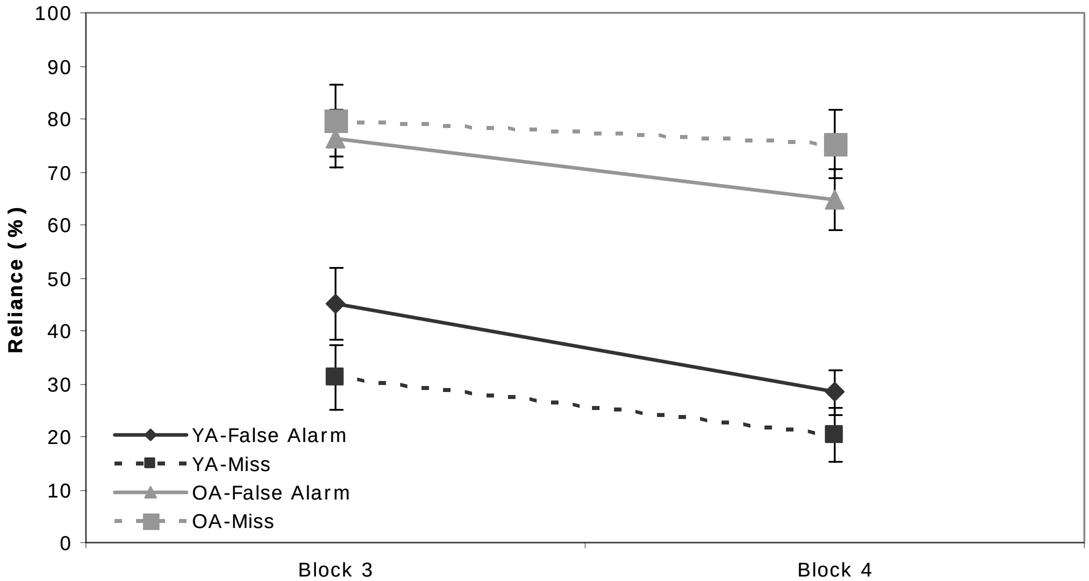


Figure 11. Reliance behavior by younger and older adults by Error Type in Block 3 and Block 4. The Block by Error Type is statistically significant for older adults.

Expectancy, error type, and compliance on the automation during blocks 1, 2, & 3. Compliance behavior is presented in Figure 12 and represents checking on the automation when the automation was in the alarm state. Older adults complied with the automation 72.3 percent of the time collapsed across Error Type, Expectancy, and Block whereas younger adults only complied with the automation 55.2 percent of the time ($F(1, 108) = 11.49$, $p < .05$, $\eta^2 = .096$). Both groups complied with the automation well below the actual reliability of the system (90%). As expected, there was a statistically significant main effect of Error Type. Previous research has suggested that false alarms reduce compliance whereas misses do not (e.g., Dixon et al., 2007). This finding was supported as participants in the false alarm condition complied with the automation 48.5 percent of the time while participants in the miss condition complied with the automation 78.9 percent of the time ($F(1, 108) = 36.10$, $p < .05$, $\eta^2 = .251$). There was a significant

Block by Error interaction ($F(2, 107) = 6.75$, $p < .05$, $\eta^2 = .112$). There was no change in compliance behavior from Block 1 to Block 3 for participants in the miss condition ($t(59) = 1.24$, $p > .0056$). In contrast, participants in the false alarm condition significantly reduced their compliance behavior from Block 1 to Block 3 ($t(59) = 3.39$, $p < .0056$). Expectancy did not have a significant effect on compliance behavior.

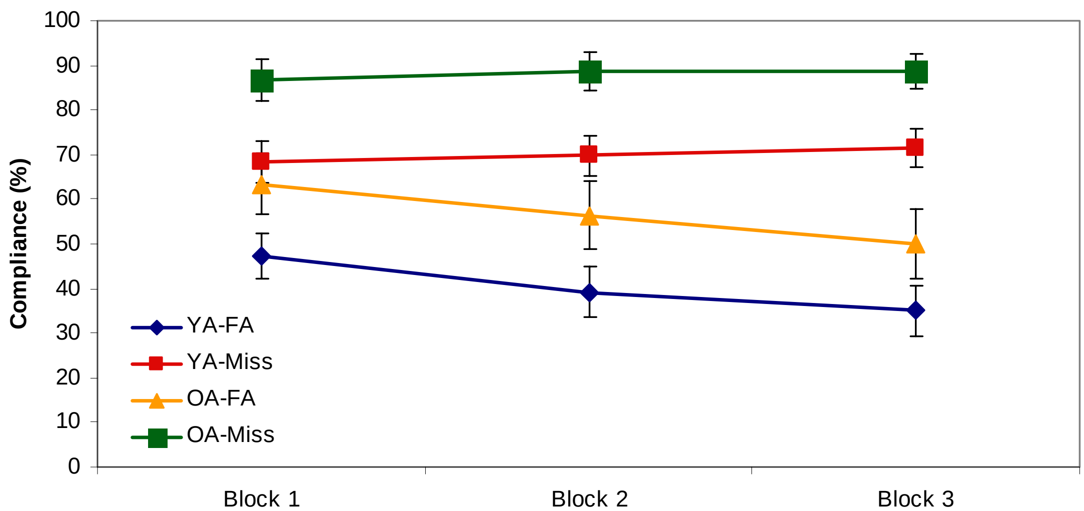


Figure 12. Compliance by Younger and Older Adults in the False Alarm and Miss conditions across Blocks 1, 2, and 3.

Expectancy, error type, and compliance on the automation: A block-by-block analysis. Compliance data for younger and older adults in each of the Expectancy and Error Type conditions are presented in Figures 13 and 14. Similarly to reliance, younger adults exhibited significant Expectancy effects during Block 1 while older adults did not. There was a great deal of variance and the only significant difference was between the High and Standard Expectancy groups, although differences between High and Low Expectancy groups were approaching significance. Again, by Block 2 the Expectancy effect had disappeared. This lends more support for the relatively short life span of Expectancy effects at least as manipulated in the present study.

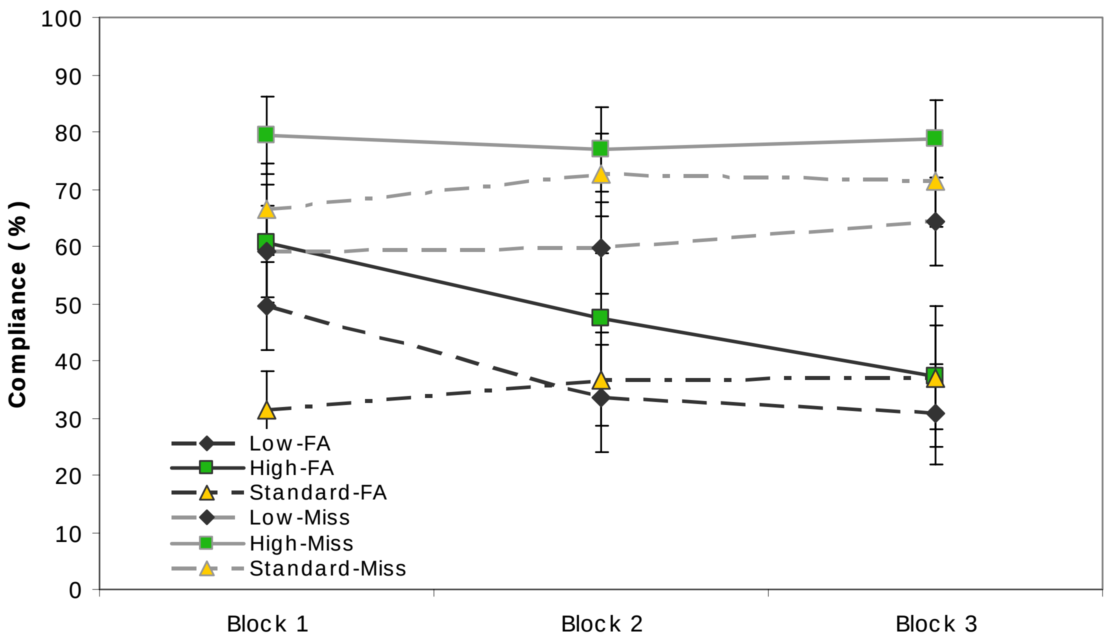


Figure 13. Compliance by Younger Adults by Expectancy (Low, High, and Standard) and Error Type (False Alarm and Miss) across Blocks 1, 2, and 3.

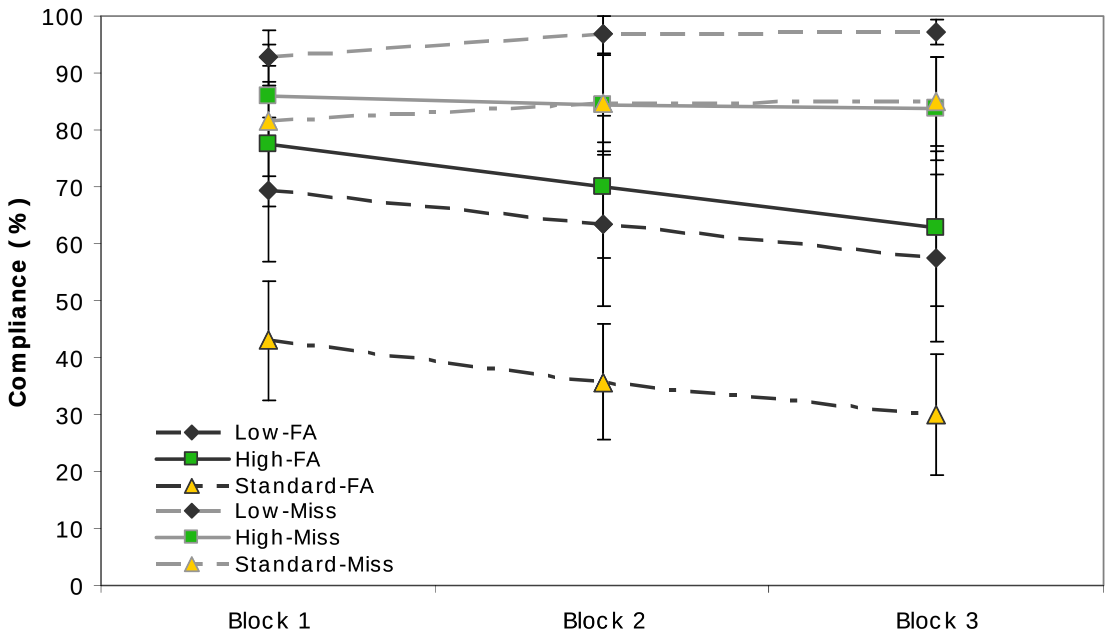


Figure 14. Compliance by Older Adults by Expectancy (Low, High, and Standard) and Error Type (False Alarm and Miss) across Blocks 1, 2, and 3.

Effect of transfer to Block 4 (60%) on compliance. Younger adults complied with the automation 53.3 percent of the time during Block 3 and 56.3 percent of the time in Block 4 ($F(1, 54) = 1.72$, $p > .05$). Older adults increased their compliance from 69.4 percent during Block 3 to 73.9 percent during Block 4 ($F(1, 54) = 10.20$, $p < .05$, $\eta^2 = .159$). For both younger ($F(1, 54) = 55.36$, $p < .05$, $\eta^2 = .506$) and older adults ($F(1, 54) = 23.62$, $p < .05$, $\eta^2 = .304$), participants in the false alarm condition complied with automation less than participants in the miss condition (See Figure 15).

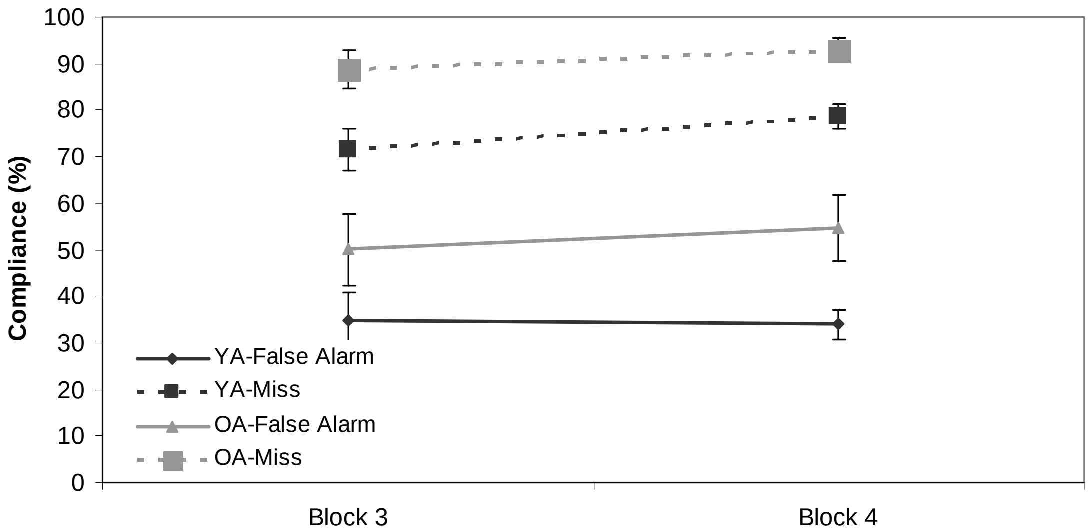


Figure 15. Compliance behavior by younger and older adults by Error Type in Block 3 and Block 4.

Summary of dependency on, reliance on, and compliance with the automation. Overall, expectancies of likely automation performance significantly influenced dependency, reliance on, and compliance with the automation for younger adults, but the effect was relatively short lived. During Block 1, younger adults in the High Expectancy condition depended on, relied on, and complied with the automation significantly more than participants in the Standard Expectancy condition and numerically more than participants in the Low Expectancy condition. Older adults did not exhibit significant expectancy effects. Participants in the false alarm condition increased reliance and reduced compliance across the first three blocks. In contrast, participants in the miss condition maintain relatively consistent reliance and compliance behavior across the first three blocks.

When younger adult participants transferred from the higher reliability automated system in Block 3 to the lower reliability automated system in Block 4, they significantly reduced their reliance regardless of Expectancy level or Error Type. However, younger adults did not change their compliance when they transferred from Block 3 to Block 4. When older adults transferred from Block 3 to Block 4, only participants in the false alarm condition significantly reduced their reliance. However, older adults increased their compliance when they transferred from Block 3 to Block 4.

3.4 Performance Dispatching Trucks and Receiving Packages

The goal of introducing automation is to make tasks easier, more efficient, or safer for humans (Dzindolet et al., 2003). Ultimately, it is the overall task, as well as subtask performance that should be of greatest concern to human-automation researchers. Simply introducing automation does not guarantee superior task performance; a number of variables can differentially affect primary tasks (those that are being aided by the automation) and secondary tasks (those that are not being aided by the automation). The following section will present performance results for the dispatching trucks task (the primary task) followed by results for the receiving packages task (secondary task). In both tasks, performance was calculated as the proportion of correct responses.

Correctly dispatched trucks (Blocks 1-3). Dispatching trucks was the primary task in this experimental procedure and was supported by the AWMS. As discussed in the method section, the AWMS performed at 90 percent reliability for the first three blocks. There was a significant main effect of age ($F(1, 108) = 45.28$, $p < .05$, $\eta^2 = .295$). Collapsed across Block, Error Type, and Expectancy, younger adults correctly dispatched trucks 95.5 percent of the time whereas older adults correctly dispatched trucks only 85 percent of the time. Younger adults performed at a higher level than had they completely depended on the automation. In contrast, older adults performed worse than they would have if they had totally depended on the automation. Figure 16 illustrates the main effect of age and block.

There was a significant main effect of block ($F(2, 107) = 8.74$, $p < .05$, $\eta^2 = .075$). Across the first, second, and third Block, average performance increased from 88.2 to 90.5 to 91.9 percent respectively. Paired sample t-tests revealed that performance increased from Block 1 to Block 3 ($t(119) = 3.73$, $p < .0167$). There were no significant main effects for Expectancy, $F(2, 108) = .26$, $p > .05$, or Error Type, $F(1, 108) = .53$, $p > .05$.

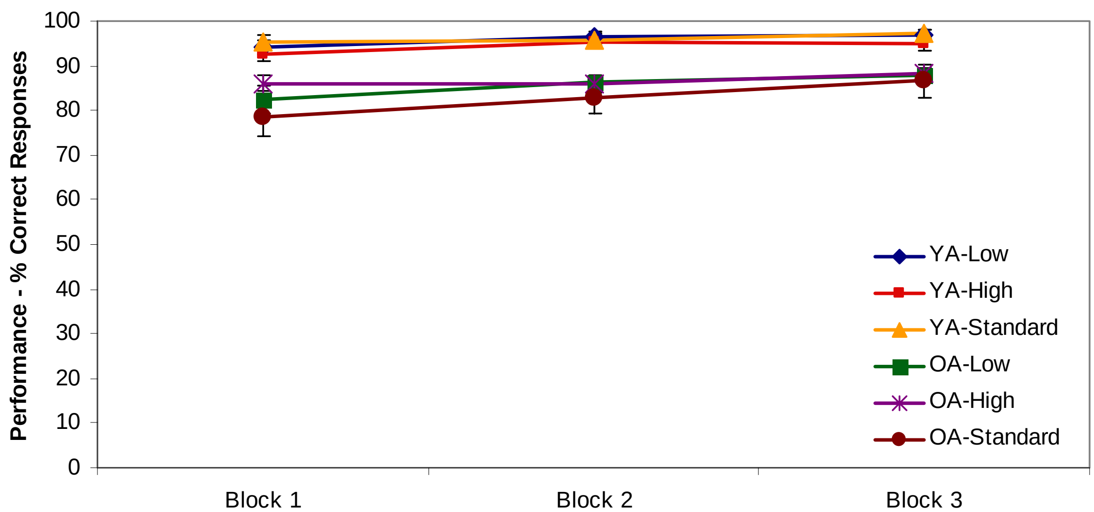


Figure 16. Mean percentage of correctly dispatched trucks by younger and older adults in the Low, High, and Standard Expectancy conditions across the first three Blocks.

Correctly received packages (blocks 1-3). Receiving packages was the secondary task in the current study. Participants were not aided on the receiving task by any automated system. Performance on the receiving task was measured as the percentage of correctly received packages as a function of the total number of packages. These data are shown in Figure 17.

Overall, younger adults outperformed older adults on the receiving task, correctly receiving 96.4 percent of all packages collapsed across the first three blocks compared to 91.8 percent respectively. There were significant main effects of Age, $F(1, 108) = 18.89$, $p < .05$, $\eta^2 = .149$ and Block, $F(2, 107) = 17.80$, $p < .05$, $\eta^2 = .250$, where participants significantly improved their performance from Block 1 to Block 3, $t(2) = 1.75$, $p < .0167$. The Age by Block interaction was significant, $F(2, 107) = 18.60$, $p < .05$, $\eta^2 = .258$ due to the older, but not younger, adults significantly improving performance from Block 1 to Block 3, $t(59) = 5.92$, $p < .05$. During the first Block younger adults outperformed older adults, correctly receiving 96.5 and 89.5 percent of packages respectively, $t(118) = 5.38$, $p < .05$, but by Block 3 there was no statistical difference in performance between younger (96%) and older (93.5%) adults, $t(118) = 2.47$, $p < .05$.

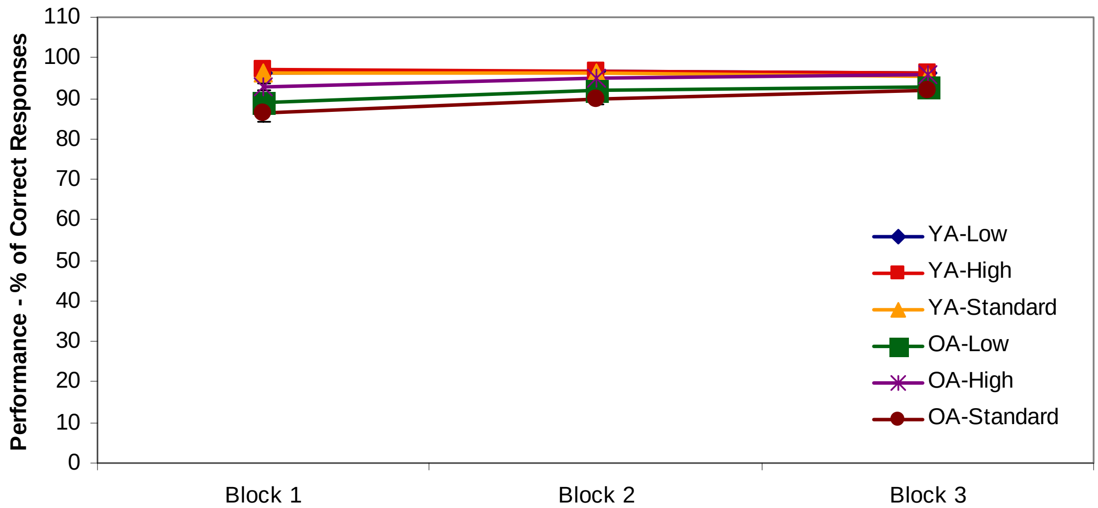


Figure 17. Percentage of correctly received packages by younger and older adults in the Low, High, and Standard Expectancy conditions across the first three Blocks.

3.5 Performance Transfer from Block 3 to Block 4

The purpose of transferring to the less reliable automated system in Block 4 was to help gain insight into what was learned about the automated system during the first three blocks. In the current task, the automation was 90 percent reliable during the first three Blocks and 60 percent reliable during Block 4. Comparisons were made between the third Block, when participants were the most practiced, and Block 4.

Performance dispatching trucks during transfer to the less reliable automation. The data are presented in Figure 18. Overall, participants correctly dispatched 92 percent of the trucks during Block 3 but only correctly dispatched 79.1 percent of the trucks during Block 4 ($F(1, 108) = 151.78$, $p < .05$, $\eta^2 = .584$). Older adults correctly dispatched 87.5 percent of trucks during Block 3 and only 66.7 percent during block 4. Younger adults correctly dispatched 96.3 percent of trucks during Block 3 and 91.6 percent during Block 4. This Block by Age interaction was statistically significant ($F(1, 108) = 60.10$, $p < .05$, $\eta^2 = .358$). Performance by young adults was better than older adults during both blocks (block 3: $t(118) = 5.67$, $p < .0125$; transfer block: $t(118) = 8.05$, $p < .0125$). In addition, both younger and older adults' performance decrease significantly from Block 3 to Block 4, but the decrease for younger adults was smaller ($t(59) = 4.28$, $p < .0125$) compared to the decrease for older adults ($t(59) = 10.78$, $p < .0125$). Correct truck dispatching by older adults during Block 3 (87.54%) and Block 4 (66.67%) is close to the actual system reliability during those same blocks (90% and 60% respectively) suggesting a close performance and automation linkage. One possibility for this finding is that older adults never learned how the automation functioned and thus how to mitigate automation errors. A second possibility is that the overall task workload was high for older adults necessitating high dependence on the automation.

One could argue that older adults placed more emphasis on the secondary task and essentially allowed the automation to take care of the primary task. Results from the exit interview revealed that all participants reported attending more to the receiving task than the dispatching task. There was no significant difference between younger and older adults self-report of the task on which they placed the most emphasis. Older and younger adults could have different performance-resource functions (Norman & Bobrow, 1976). It would follow that older adults' actual emphasis on the receiving task resulted in little "spare capacity" for the dispatching task while younger adults had a "reserve capacity" that could be shifted to the dispatching task.

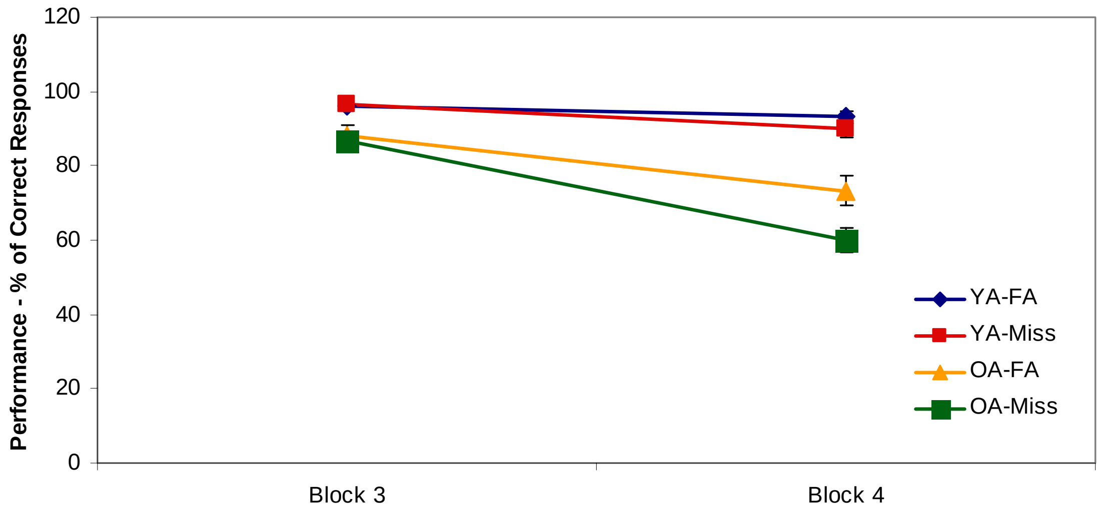


Figure 18. Performance dispatching trucks by younger and older adults in the False Alarm and Miss conditions.

There was a significant Block by Error Type interaction ($F(1, 108) = 12.29$, $p < .05$, $\eta^2 = .102$). Statistically, there were no differences between performance by participants in the false alarm condition and the miss condition in Block 3 or in Block 4 (Block 3: $t(118) = .41$, $p > .0125$; Block 4: $t(118) = 2.12$, $p > .0125$). There were significant decreases in performance from block 3 to the transfer block for participants in the false alarm condition ($t(59) = 5.80$, $p < .0125$) and in the miss condition ($t(59) = 8.00$, $p < .0125$), but the decrease was greater for participants in the miss condition.

Performance receiving packages during transfer to the less reliable automation. Receiving packages showed no negative effect of transferring to the 60 percent reliability block. Overall, there were no significant effects of Age, Error Type, or Expectancy. This result is consistent with the emphasis reports mentioned above and, for older adults, consistent with the larger drop in performance during the transfer to a lower automation accuracy.

Summary of performance dispatching trucks and receiving packages. Overall, younger adults dispatched a greater percentage of trucks compared to younger adults across all blocks. When participants transferred to the less reliable automation, both younger and older adults' performance dispatching trucks decreased, however, the performance decrease for older adults was much greater than that for older adults. In fact, older adults' performance dispatching trucks matched closely to the actual reliability of the system.

Younger adults received packages at a very high level throughout the experiment. Older adults increased performance across the first three blocks to a point that was statistically equivalent to younger adult performance. When transferred to the less reliable automation, both younger and older adults maintained a high level of performance and did not appear to be affected by the change in automation performance.

3.6 Analysis of Trust: General & Specific

The first analysis regressed the general trust scores on the specific trust scores. If the general trust scores significantly predicted specific trust, then it could be argued that any differences in specific trust scores could be due to initial general trust. A linear regression was conducted using values on the general trust questionnaire as independent variables. The specific trust score was used as the dependent variable. The R^2 value was very small ($R^2 = .048$) and statistically non-significant ($F(4, 115) = 1.44$, $p > .05$) indicating that differences in specific trust must be due to the experimental manipulations or experience with the automation and not initial bias.

Trust was measured with questions that used a Likert-like response scale. The data therefore are ordinal, however, for the purpose of the current investigation, the Likert-like response scale was assumed interval.

Analysis of general trust. The general trust questionnaire was one of the first things participants completed in the experiment. They were provided with a very general, high level definition of automation and then completed the questionnaire. The purpose of the general trust questionnaire was two-fold: (1) to determine if participants entered the experiment with equivalent trust levels and (2) to establish a baseline trust level for purpose of statistical comparison with trust measures to be administered later in the experimental protocol. The general trust data are presented in Table 7.

Table 7
General Trust Scores of Younger and Older Adult Participants in all 6 Between Subject Groups.

		Younger Adults		Older Adults	
		Mean	SD	Mean	SD
Low					
	False Alarm	5.30	1.25	4.50	1.08
	Miss	5.30	0.82	5.30	0.95
High					
	False Alarm	4.80	0.92	4.70	1.25
	Miss	5.30	0.48	4.90	1.20
Standard					
	False Alarm	4.90	0.88	5.20	2.15
	Miss	5.40	0.97	4.50	1.18

Note: Response scale was from 1 to 7 where 7 indicated complete trust in the automation and 1 indicated no trust in the automation.

There were no significant main effects of Age, Expectancy, or Error Type indicating that prior to being exposed to the experimental manipulations, participants had statistically equivalent levels of trust in automation (see Figure 19). The null effect of Expectancy and Error Type was anticipated given that participants were randomly assigned to the between subjects groups. It is possible, given the vague definition of automation provided to participants, that younger and older adults had different mental representations of what constituted an automated system, however, there is no way to tell whether that was the case here.

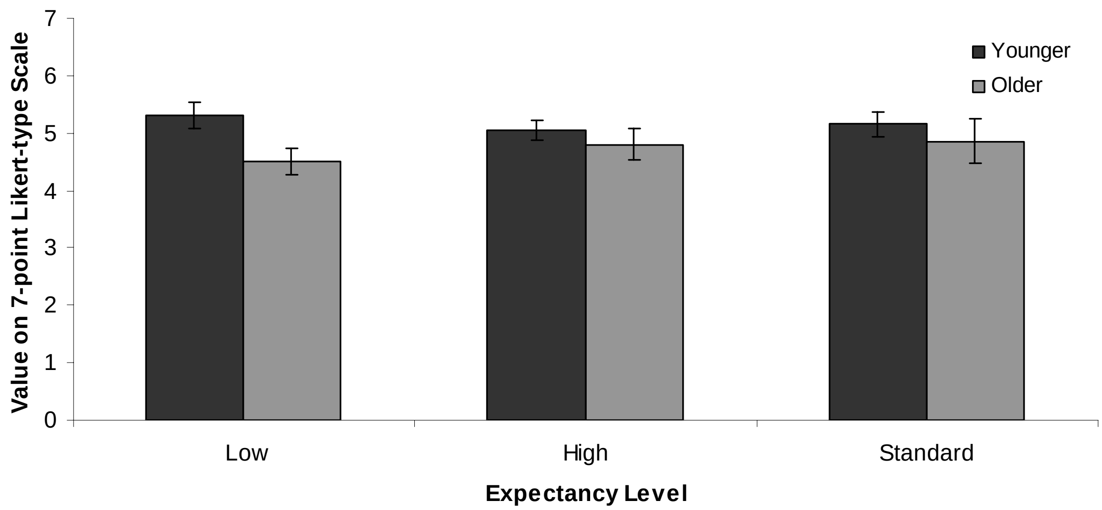


Figure 19. General trust of younger and older adult participants by expectancy condition.

Analysis of specific trust. As previously shown, participants entered the experiment with comparable levels of trust in automation. Of interest was whether there would be differences in trust after interacting with the automation that performed at 90 percent reliability. It is important to emphasize that, regardless of level of expectancy; all participants interacted with the identical system. The only difference was between the false alarm and miss conditions, the difference being the type of error, not the overall reliability of the system.

No specific predictions were made regarding the effects of the experimental manipulations on specific trust. It has been suggested that expectations are the building blocks of trust (Barber, 1985), so one proposed possibility, following from Barber's proposal, was that participants in the high expectancy group would have higher specific trust compared to participants in the low expectancy group. There was a significant Age difference in trust in the system. Older adults reported higher trust in the automation compared to younger adults ($F(1, 108) = 26.51$, $p < .05$, $\eta^2 = .197$). This is possibly because older adults relied on and complied with the automation at a higher level than younger adults and thus to avoid cognitive dissonance reported higher levels of trust in the system. Another possibility is that the tasks were much more difficult for older adults compared to younger adults and in combination with the relatively high level of system reliability, led older adults to trust the automation more than younger adults.

There was also a significant Age by Expectancy interaction ($F(2, 108) = 4.03$, $p < .05$, $\eta^2 = .069$). Further analyses revealed that there were no significant differences in specific trust for young adults between Expectancy conditions. In contrast, there were significant differences in specific trust for older adults between Expectancy conditions. Specifically, older adults in the High Expectancy condition had higher trust in the AWMS than older adults in the Low Expectancy condition ($t(19) = .80$, $p < .05$) and than older adults in the Standard Expectancy condition ($t(19) = .60$, $p < .05$) (See Figure 20). These findings, at least for the older adults, lend support for Barber's (1985) claim that expectancies help determine trust. However, older adults did not rely on and comply with the automation differentially between Expectancy conditions. Perhaps if the task had been easier for older adults they would have shown expectancy effects in behavioral reliance and compliance similarly to younger adults.

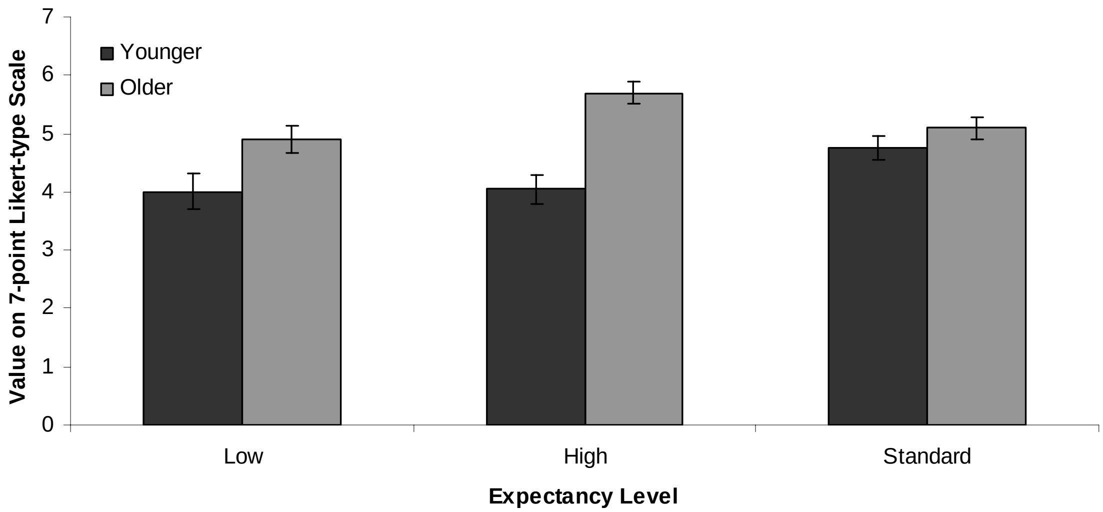


Figure 20. Interaction between Age and Expectancy condition on specific trust.

CHAPTER 4

DISCUSSION

There were several findings of note from the current study. Firstly, significant Expectancy effects were found for dependence, reliance, and compliance but only for younger adults and only during the first block. By the second and third blocks, the expectancy effects had all but disappeared. Second, during transfer of performance on the primary (truck dispatching) task, younger adults maintained a high level of performance whereas older adult performance decreased significantly, matching closely the performance of the automated system. This can be explained nicely by the transfer of dependence finding, where younger adults significantly decreased dependence during Block 4 but older adults did not. Third, Error Type had no significant effect on reliance behavior but exerted a fairly strong statistically significant effect on compliance behavior, replicating findings by Dixon et al. (2007). Finally, despite comparable levels of general trust, older adults reported higher trust in the AWMS than younger adults.

4.1 Expectancy Effects and Dependence, Reliance, and Compliance

The primary thrust of the present study was to gain an understanding of the influence operator expectations have on dependence, reliance, and compliance. Expectations were operationally defined as participants' belief regarding the automation's likely performance during the experiment. There were two primary motivations for the research: Previous research on human-automation interaction has traditionally provided participants with knowledge of system errors, misses and false alarms, which may bias participant expectations and subsequent behavior. In addition, more and more often people from all age groups find themselves in situations where they are forced to interact with unfamiliar automated systems of which they have no level of knowledge regarding errors, performance, and functioning. It is imperative therefore that we begin to understand user interactions with automation in such situations.

In preliminary research, findings suggested that high and low expectancies had a strong effect on dependence in a younger adult population, increasing and reducing dependence respectively (Mayer et al., 2006). The preliminary study used only a 30-minute experimental block, so goals for the current study were to see if (a) the expectancy effect could be replicated and if (b) the expectancy effect extended beyond 30 minutes. In addition, age-related effects were investigated since older adults stand to benefit greatly from automation and age-related investigations of expectancy effects are missing in the literature.

The results show that the current study successfully replicated the preliminary study. Younger adults in the high expectancy condition, during the first block, depended on the automation more than participants in the low and standard conditions. These differences were achieved using a fairly subtle manipulation of the automation description. These results differ from Madhavan and Wiegmann (2007) who did not find any expectancy effects for 90 percent reliable systems. One explanation for the difference is that the Madhavan and Wiegmann task was a binary detection luggage screening task where participants had to simply decide to stop or pass a piece of luggage. The relatively simple nature of the task may have allowed participants more attention for recognizing the high reliability of the automation. In contrast, the dual-task used in the current study was high paced and highly demanding which may have prevented participants from accruing sufficient knowledge of system reliability requiring them to rely on their expectancies.

The effect of expectancy was gone by block 2 suggesting that expectancy effects are short lived and fail to extend beyond a 20-30 minute time period. Madhavan and Wiegmann (2007) found a similar result for a 70 percent reliable system and attributed it to a violation of the “perfect automation schema” originally proposed by Dzindolet et al. (2001). When participants interacted with an expert automated system, they dramatically reduced their reliance and compliance in the face of automation errors (Madhavan and Wiegmann). This may indeed be the case except that in the current study, participants’ expectations were manipulated to shift participants in the low expectancy condition away from a “perfect automation schema.” It should be stressed that a short-lived effect of expectancy should not be equated with an unimportant effect of expectancy. In certain domains even a single error can be catastrophic.

Older adults did not exhibit statistically significant expectancy effects. However, numerically, there were differences in dependence that were similar to younger adults. Overall, there was a great deal of variance in the data particularly with the older adults. The likely source of the variance was the vague understanding participants had regarding the functioning of the automation. As mentioned previously, most automation studies provide participants with knowledge of system functioning in the form of either the type of errors the automation commits, the overall reliability of the system, or a combination of the two, which plausibly leads to less variability. The current study was specifically designed to ecologically sample the uncertainty people have when interacting with unfamiliar systems, so the system descriptions were purposely vague. The potential downside of this approach, however, is that people may develop an understanding of system performance in different ways and at different rates which results in higher variance.

One advantage of providing participants with explicit information related to system errors and system reliability is that it can reduce variance. However, are participants being biased to depend less on the automation because of this information? Results from the current study suggest that yes, providing participants with some knowledge of system performance does degrade their overall dependence. Both younger and older adults in the standard expectancy condition depended, relied, and complied with the automation at a lower rate than participants in the high and low expectancy conditions, despite reporting equivalent expectancies to the high expectancy condition on the expectancy questionnaire. One explanation is that because participants in the standard condition had some understanding of system errors, they were better able to recognize and encode automation errors, resulting in an underestimation of automation reliability.

4.2 Task Performance Dispatching Trucks and Receiving Packages

Performance is an important variable because it is performance on the automated task and any secondary tasks that will dictate whether the human-automation interaction is a failure or a success. In the current study, performance was measured as the percentage of correct responses given on the primary task (the truck dispatching task) and the secondary task (the receiving packages task).

Overall, younger adults outperformed older adults on both tasks. This was not surprising given the wealth of evidence showing younger adults outperforming older adults in dual-task situations (e.g., Rogers et al., 1994). However, for the receiving packages task, older adults gradually improved their performance such that by the third block their performance was statistically equivalent to that of younger adults. In addition, both younger and older adults maintained a high level of performance when transferred to the less reliable automation in Block 4 on the receiving packages task. This was a little surprising. It would be expected that if participants really were attending to both tasks, then performance on the secondary task would decline when the reliability of the automation declined during the transfer block because more attention was required to compensate for the poorly performing automation.

There are several possible reasons for why this was not the case in this study. The most likely possibility is that participants chose to focus the majority of their attention on the secondary task, depending more on the automation, during the transfer block. If this was the case, then one would expect performance on the primary task to match closely to the actual performance of the automation. This did not seem to be the case for younger adults, as their performance stayed high on both tasks and their dependence behavior changed as the reliability of the system changed.

It seemed, however, to be the case that older adults focused more attention on the secondary task. Their performance dispatching trucks during the third block (87.54%) was roughly equivalent to the performance of the automation (90%) during the third block. In addition, their performance during the transfer block (66.67%) was roughly equivalent to the performance of the automation (60%) during the transfer block. This can be explained by older adults' high level of dependence behavior (relative to younger adults) that did not change from block 3 to transfer suggesting that they may have allowed the automation to take more responsibility for the primary task so they could focus more attention on the secondary task.

Dual-task scenarios are difficult, particularly for older adults (e.g., Rogers, Bertus, & Gilbert, 1994) and this may be what happened in the current study. Because the task was so difficult, and because they had help on the primary task, they focused on the secondary task. In the exit interview, participants were asked if they paid more attention to one task compared to the other. On average, both younger and older adults reported paying more attention to the receiving packages task, but there was no significant difference between younger and older adults. This is likely because the receiving packages task was a continuous task that did indeed require more attention as a whole.

Another explanation is that older adults may be more averse to risk taking and more conservative in their responses in an attempt to reduce errors. Older adults may view the automation as being responsible for the primary task of dispatching trucks and take more personal responsibility for the secondary task of receiving packages.

4.3 Error Type

There has been debate in the literature regarding whether system misses or false alarms are more detrimental to human-automation interaction (e.g., Johnson, 2004; Meyer, 2004; Dixon, Wickens, & McCarley, 2007). More recently, evidence suggests that, from the perspective of performance measures, misses and false alarms have differential effects on primary and secondary task performance (e.g., Dixon & Wickens, 2006; Dixon, Wickens, & McCarley). The findings are such that false alarms are more detrimental to the primary task while misses are more detrimental to the secondary task.

The current research failed to replicate the finding. During the first three blocks, there was no significant effect of error type on dispatching truck performance. There was a significant effect during the transfer block, but it was contrary to the previous literature. Performance on the primary task by participants in the miss condition decreased more than performance by participants in the false alarm condition. In the secondary task, there were no effects of error type during the first three blocks or during transfer. This is also contrary to previous findings in the literature. One possible explanation is that false alarms provide more information compared to misses (Madhavan, Wiegmann, & Lacson, 2006). It is plausible that participants in the false alarm condition maintain a higher level of performance because they have a more salient cue (i.e., the automation alert) that allows them to better manage their attention between the two tasks. Participants in the miss condition suffer more when they transfer to the less reliable automation because they do not have access to a salient cue that can help them balance their shifts in attention.

Across the first three blocks, younger and older adults in the miss condition maintained a constant level of both reliance and compliance. Adjustments were made by participants in the false alarm condition who decreased compliance and increased reliance across the first three blocks. It appears as though participants who interact with automation that misses set a level of reliance and compliance and maintain that level so long as the automation's reliability remains constant. Sanchez (2006) found that participants in a miss condition required more time to appropriately adjust their behavior. It is possible that given more time, participants in the miss condition of the current study would have adjusted their behavior.

There is also evidence in the literature that false alarms and misses have differential effects on reliance and compliance. As outlined in the introduction, originally, misses were thought to result in lower reliance while false alarms were thought to result in lower compliance (Meyer, 2001; 2004). More recently, evidence has suggested that misses result in lower reliance while false alarms result in lower reliance and compliance (Dixon, Wickens, & McCarley, 2007).

When transferred to the less reliable automation, younger adults in the miss condition significantly reduced their reliance but not their compliance behavior. Younger adults in the false alarm condition reduced both their reliance and compliance. A possible reason that false alarms have such broad effects is that they are more salient and thus more memorable (Madhavan, Wiegmann, & Lacson, 2006). The effects of false alarms seem to creep into reliance. This 'reliance-creep' has now been demonstrated in a couple of different studies using unique paradigms.

4.4 Trust

Trust is an important variable that has been shown to influence dependence on automation when a thorough understanding of the system is unavailable or if the system functions are too complex (Lee & See, 2004). In addition, trust has been thought to be influenced by expectancies (Barber, 1985) although no empirical studies were found during the literature review. Thus, no specific hypotheses were made regarding the subjective measure of trust.

Overall, younger and older adults entered the experiment with comparable trust in automated systems. Congruent with findings by Dzindolet et al. (2001), participants reported having above neutral feelings of trust in automation. Dzindolet et al. proposed a perfect automation schema suggesting that, ceteris paribus, people have a high level of trust in automated systems.

The specific trust questionnaire was given after exposure to the third experimental block. There was a significant age by expectancy interaction that showed that older adults in the high expectancy condition trusted the system more than older adults in the low or standard conditions. There were no significant differences between the low, high, and standard conditions for younger adults. In addition, there was no significant difference between younger and older adults in the low and standard conditions, but older adults in the high expectancy condition trusted the system more than younger adults in the high expectancy condition.

The question is why did older adults in the high expectancy condition report higher trust in the automation than participants in any of the other expectancy conditions? It is difficult to speculate about the answer to this question. Analyses were conducted on ordinal data assumed to be interval so the findings need to be taken with caution. In addition, the specific trust questionnaire has not been validated. If the behavioral data had matched with the trust score then there would be more inclination to believe findings from the trust data.

4.5 Future Directions

The current study added to the understanding of the role of expectancies on dependence, reliance, and compliance. In addition, a better understanding of the effects of type of automation errors was established from this study. However, there are still unanswered questions related to both the effects of user expectancies and automation errors on dependence, reliance, compliance, and ultimate task performance.

Older adults had, relative to younger adults, high dependency, reliance on, and compliance with the automation. One proposed explanation is high workload for older adults. However, there is no empirical data that show that the reason for older adults' levels of dependency, reliance on, and compliance with automation is due strictly to workload. Research needs to be conducted to determine if indeed this is the case or if the high levels are due to some other factor such as task responsibility.

The development of automation usage patterns is not well understood. The current study showed that expectations initially influence behavior, but the mechanism of dependence, reliance, and compliance is unknown. It appears as though people begin to base their behavior on some understanding of the system that accrues with experience but the nature of the information, the development and progression of this learning is still not understood. It is unclear if people begin by basing their behavior on the vague expectations of the automation and then progress to a general but more concrete understanding that may include general system reliability and then finally progress to a more specific understanding that may include knowledge of the specific criterion of the automation. It is also unclear what the mechanism for such a progression would be.

There are still open questions surrounding the effects of error type on both dependence behavior and performance. More and more evidence, the current study included, suggests that false alarms provide more information to users. The information does not appear to be complete and lacks sensitivity, but it does, particularly for older adults, provide a basic understanding that the automation is making an error and that it generally happens during the alarm state. However, it is unknown exactly what it is about false alarms that provide the added information. One possibility is that it is the salience of false alarms that make them more informative. Although, it could be the mere fact that false alarms only occur during the alarm state of the automation and that people consider automation to be functional or active during the alarm state that drives the informative nature of false alarms, but this remains an open question.

CHAPTER 5

CONCLUSION

The way people interact with automation is dependent on both the expectations they have regarding the likely performance of the system and the actual performance of the system. Although expectancy effects appear to be short lived, in certain domains the cost of committing an error because of inappropriate expectations may be catastrophic. Older adults appear to be slightly more resistant to expectancy effects but their overall understanding of automation functioning appears to lag behind that of younger adults. There appears to be a differential effect of automation error type. False alarms appear to provide added information allowing people to more appropriately calibrate their reliance and compliance behavior when the automation is of a consistent reliability. But when people transfer to a less reliable system false alarms appear to be more detrimental than misses, reducing both reliance and compliance behavior suggesting that the sensitivity of the information provided by false alarms is low.

APPENDIX A

GENERAL TRUST QUESTIONNAIRE

An automated system is a technologically-based system used to partially or fully assist the human in tasks involving sensing, detecting, information processing, making decisions and/or executing actions.

Please circle the number that best describes your feeling, impression, or belief.

1. Automated systems are deceptive

1 2 3 4 5 6 7

Not at All **Extremely**

2. Automated systems behave in an underhanded manner

1 2 3 4 5 6 7

Not at All **Extremely**

3. I am suspicious of automated systems' intent, action, or outputs

1 2 3 4 5 6 7

Not at All **Extremely**

4. I am wary of automated systems

1 2 3 4 5 6 7

Not at All **Extremely**

5. Automated systems' actions have a harmful or injurious outcome

1 2 3 4 5 6 7

Not at All **Extremely**

6. I am confident in automated systems

1	**2**	**3**	**4**	**5**	**6**	**7**

Not at All **Extremely**

7. Automated systems provide security

1	**2**	**3**	**4**	**5**	**6**	**7**

Not at All **Extremely**

8. Automated systems have integrity

1	**2**	**3**	**4**	**5**	**6**	**7**

Not at All **Extremely**

9. Automated systems are dependable

1	**2**	**3**	**4**	**5**	**6**	**7**

Not at All **Extremely**

10. Automated systems are reliable

1	**2**	**3**	**4**	**5**	**6**	**7**

Not at All **Extremely**

11. I can trust automated systems

1	**2**	**3**	**4**	**5**	**6**	**7**

Not at All **Extremely**

12. I am familiar with automated systems

1	**2**	**3**	**4**	**5**	**6**	**7**

Not at All **Extremely**

13. To what extent do you think you could count on an Automated System to do its job?

1	2	3	4	5	6	7

Not at All **Completely**

14. Overall, how much would you trust an Automated System?

1	2	3	4	5	6	7

Not at All **Completely**

15. Please indicate how often you think an Automated System would provide correct information (using a %).

(Example: I think an Automated System would provide correct information ##% of the time)

_______%

SPECIFIC TRUST QUESTIONNAIRE

Please answer the following questions about the Automated Warehouse Management System with which you just interacted.

1. How much do you trust the Automated Warehouse Management System now that you have interacted with it?

1	2	3	4	5	6	7
Not at all						**Completely**

2. Please indicate how often you believe the Automated Warehouse Management System provided correct information (using a %).

   (Example: I think the Automated Warehouse Management System provided correct information ##% of the time)

   ______%

3. Please indicate how much you relied on the Automated System (using a %).

   (Example: I relied on the Automated System ##% of the time)

   _____%

4. To what extent can you count on the Automated Warehouse Management System to do its job?

1	2	3	4	5	6	7
Not at all						**Completely**

5. Please indicate the reliability of the Automated Warehouse Management System (using a %)
   (Example: I think the Automated Warehouse Management System was ##% reliable)

_______%

6. My performance in this task would have been better without the Automated Warehouse Management System

1	2	3	4	5	6	7
Agree						**Disagree**

7. Please indicate the total number of times you believe that the Automated Warehouse Management System caused you to do the following:

a) it caused me to overload a truck _____ number of times

b) it caused me to dispatch a truck that was not full _____ number of times

APPENDIX B

EXPECTANCY DESCRIPTIONS

Low-Expectancy Description

An Automated Warehouse Management System is a system that scans the inside of truck trailers, calculates the amount of space available in the truck, loads shipments onto the truck, determines if the truck is full, and when the truck is full notifies the Warehouse Manager to dispatch the truck. The Automated Warehouse Management System that is being tested today is designed for large, high traffic commercial warehouse operations

SRT-1 Automated Warehouse Management System

We are working with a company on issues of automation, as well as being funded by the National Institute of Health for this work. Let me tell you a little about the system you will be helping us test. The company first became involved in sensory technologies in 2000 with the sole mission of creating advanced scanning and decision making systems for warehouse loading and shipping applications. In 2001, the company proposed an Automated Warehouse Management System called the SRT and in 2004 proposed a Smart Automated Warehouse Management System, the SRT -1. The company's first prototype system, the SRT-1, utilizes advanced decision algorithms and sensing technologies that have the ability to adjust to differing warehouse and loading conditions. Testing of the SRT-1 has not begun so designers are unsure of the accuracy, reliability, and robustness of the Automated Warehouse Shipping system and how it will compare to the industry standard. Because this is a first prototype Automated Warehouse Management System, it is expected that the SRT-1 will perform at a low level with some performance errors.

High Expectancy Condition

An Automated Warehouse Management System is a system that scans the inside of truck trailers, calculates the amount of space available in the truck, loads shipments onto the truck, determines if the truck is full, and when the truck is full notifies the Warehouse Manager to dispatch the truck. The Automated Warehouse Management System that is being tested today is designed for large, high traffic commercial warehouse operations

SRT-2 Automated Warehouse Management System

We are working with a company on issues of automation, as well as being funded by the National Institute of Health for this work. Let me tell you a little about the system you will be helping us test. The company first became involved in sensory technologies in 1975 with the sole mission of creating advanced scanning and decision making systems for warehouse loading and shipping applications. In 1985, the company released an Automated Warehouse Management System called the SRT and in 1997, released a Smart Automated Warehouse Management System, the SRT-1. The company's latest groundbreaking system, the SRT-2, utilizes advanced decision algorithms and sensing technologies that have the ability to adjust to differing warehouse and loading conditions. Testing of the SRT-2 indicates that it is the industry standard for accuracy, reliability, and robustness and is still considered the leader in Automated Warehouse Management systems. Because this is a well proven Automated Warehouse Management System, it is expected that the SRT-2 will perform at a high level with no performance errors.

Standard Expectancy Condition

An Automated Warehouse Management System is a system that scans the inside of truck trailers, calculates the amount of space available in the truck, loads shipments onto the truck, determines if the truck is full, and when the truck is full notifies the Warehouse Manager to dispatch the truck. The Automated Warehouse Management System that is being tested today is designed for large, high traffic commercial warehouse operations

SRT Automated Warehouse Management System

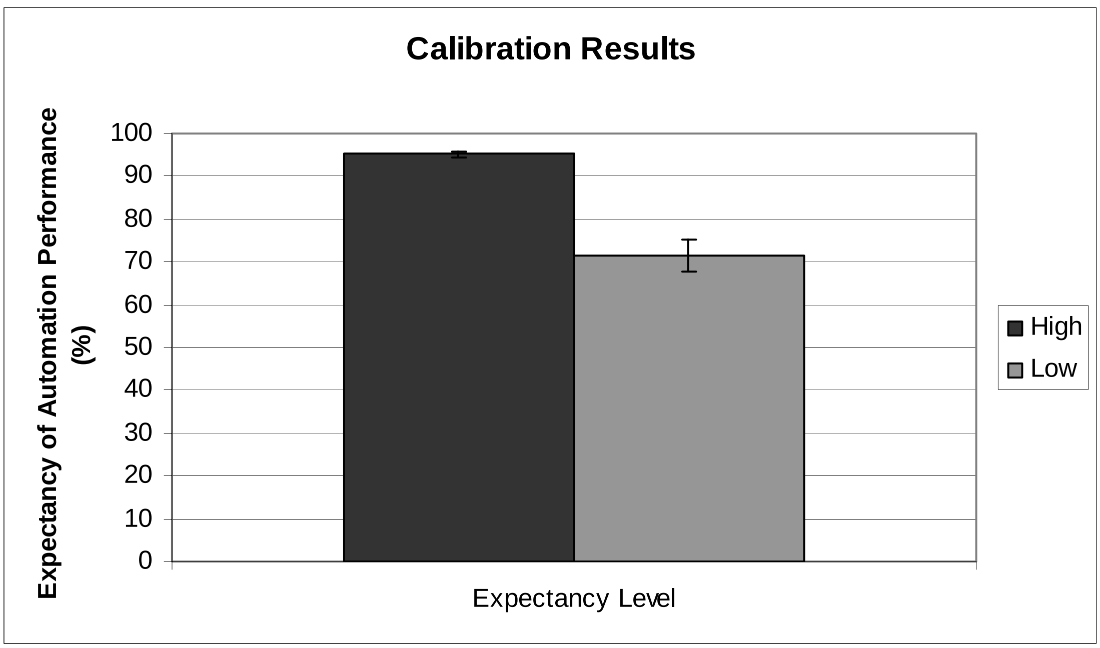
We are working with a company on issues of automation, as well as being funded by the National Institute of Health for this work. Let me tell you a little about the system you will be helping us test. The Automated Warehouse Management System that you will be interacting with today is very reliable but may make errors. Two types of errors can potentially be committed: a false alarm or a miss. A false alarm is when the system indicates that a truck is full when in fact it is not full. For example, like when smoke alarm sounds when there is no fire. A miss is when the system fails to indicate that the truck is full when in fact it IS full. For example, when there is a fire but the smoke alarm does not sound.

APPENDIX C

RESULTS OF EXPECTANCY CALLIBRATION STUDY

The expectancy descriptions for high and low expectancy conditions were calibrated and tested prior to conducting the current study. Fifty-three young adults between the ages of 18 and 28 (M = 19.27, SD = 1.27) were recruited from the Georgia Tech community. Twenty-six participants were randomly assigned to the high expectancy condition and twenty-seven participants were randomly assigned to the low expectancy condition. Participants were given the appropriate expectancy description and then were asked to complete the expectancy questionnaire.

The results indicated that the descriptions were successful at shifting expectations. On average, participants in the high expectancy condition reported that they expected the automation to provide correct information 95.17 percent of the time (SD = 3.93). In contrast, participants in the low expectancy condition reported an expectancy of correct information being provided only 71.52 percent of the time (SD = 19.56). An independent sample t-test showed that this difference was statistically significant, $t(51) = 6.05$, $p < .001$.



APPENDIX D

EXPECTANCY QUESTIONNAIRE

1. Please circle the number that corresponds to how well you expect the Automated Warehouse Management System to perform on the upcoming task.

1	2	3	4	5	6	7
Not at all well						**Perfectly**

2. Please indicate how often you believe the Automated Warehouse Management System will provide correct information (using a %). (Example: I think the Automated Warehouse Management System will be correct ##% of the time)

   _____%

3. Please indicate how much you plan to rely on the Automated Warehouse Management System (using a %).
   (Example: I plan to rely on the Automated Warehouse Management System ##% of the time)

   _____%

4. Please circle the number that corresponds to the likelihood of the Automated Warehouse Management System committing an error.

1	2	3	4	5	6	7
Not at all Likely						**Extremely Likely**

5. Please indicate how you perceive the relationship between automated systems and human users.

1	2	3	4	5
Automation works for the human		**Collaborative Team**		**Human works for the automation**

APPENDIX E

WAREHOUSE DESCRIPTION

Large Warehouse operations are high paced environments. The presence of multiple loading docks results in numerous trucks being loaded and unloaded at the same time. Full trucks are therefore often being dispatched by the minute. In addition, shipments are received and placed into inventory at an even faster pace. Warehouse managers must oversee and coordinate both the dispatching of full trucks and the receiving of shipments. The cost of dispatching trucks that are not full is enormous; likewise, the cost of overloading trucks is enormous. In large warehouses, a shipment that is placed into inventory incorrectly can be lost forever, resulting in the cost of lost inventory. To increase revenues and reduce lost shipments, warehouse owners are intensely interested in ways to improve efficiency and productivity of warehouse managers. Automated Warehouse Management Systems provide this potential.

REFERENCES

Barber, B. (1983). *The logic and limits of trust.* United States of America: Rutgers, the State University of New Jersey.

Bowers, C. A., Oser, R. L., Salas, E., & Cannon-Bowers, J. A. (1996). Team performance in automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 163-181). Mahwah, NJ: Lawrence Erlbaum.

Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology, 35,* 38-48.

Chappell, S. L. (1997). The effects of experience and automation on failure detection. *Proceedings of the 9th International Symposium on Aviation Psychology,* 773-779.

Clark, M. C., Czaja, S. J., & Weber, R. A. (1990). Older adults and daily living task profiles. *Human Factors, 32,* 537-549.

Cohen, M. S., Parasuraman, R., & Freeman, J. T. (1998). Trust in decision aids: A model and its training implications. *Proceedings, 1998 Command and Control Research and Technology Symposium,* 1-37. Department of Defense C4ISSR Cooperative Research Program.

Craik, F. I. M., & Salthouse, T. A. (Eds.). (2000). *The Handbook of Aging and Cognition* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Dixon, S. & Wickens, C. D. (2004). *Reliability in automated aids for unmanned aerial vehicle flight control: Evaluating a model of automation dependence in high workload* (Tech. Rep. No. AHFD-04-05/MAAD-04-1). Savoy, Illinois: University of Illinois at Urbana-Champaign, Aviation Human Factors Division, Institute of Aviation.

Dixon, S. R. & Wickens, C. D. (2006). Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. *Human Factors, 48,* 474-486.

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007) On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human Factors, 49,* 564-572.

Dzindolet, M. T., L. G. Pierce, et al. (2001). Predicting misuse and disuse of combat identification systems. *Military Psychology, 13,* 147-164.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies, 58*, 697-718.

Eisen, S. V., & McArthur, L. Z. (1979). Evaluating and sentencing a defendant as a function of his salience and the perceivers' set. *Personality and Social Psychology Bulletin, 5,* 48-52.

Feldman, R. S., & Theiss, A. J. (1982). The teacher and student as Pygmalions: Joint effects of teacher and student expectations. *Journal of Educational Psychology, 74*, 217-223.

Fisk, A. D., & Rogers, W. A. (2002). Psychology and aging: Enhancing the lives of an aging population. *Current Directions in Psychological Science, 11,* 107-110.

Gottlob, L. R. (2006). Age-related deficits in guided search using cues. *Psychology and Aging, 21*, 526-534.

Harris, M. J., & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin, 97,* 363-386.

Jamieson, D. W., Lydon, J. E., Stewart, G, & Zanna, M. P. (1987). Pygmalion revisited: New evidence for student expectancy effects in the classroom. *Journal of Educational Psychology, 79,* 461-466.

Jian, J., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics, 4,* 53-71.

Johnson, J. D. (2004). *Type of automation failure: The effects on trust and reliance in automation.* Unpublished master's thesis, Georgia Institute of Technology, Atlanta.

Larzelere, R. E., & Huston, T. L. (1980). The dyadic trust scale: Toward understanding interpersonal trust in close relationships. *Journal of Marriage and the Family, 42,* 595-603.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics, 35,* 1243-1270.

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies, 40,* 153-184.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*, 50-80.

Madhavan, P., & Wiegmann, D. A. (2005). Effects of information source, pedigree, and reliability on operators' utilization of diagnostic advice. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting* (487-491). Santa Monica, CA: Human Factors and Ergonomics Society.

Madhavan, P. & Wiegmann, D. A. (2007). Effects on information source, pedigree, and reliability on operator interaction with decision support systems. *Human Factors, 49,* 773-785.

Madhavan, P., Wiegmann, D. A., & Lacson, F. C. (2006). Automation failures on tasks easily performed by operators undermine trust in automated aids. *Human Factors, 48,* 241-256.

Mayer, A. K., Sanchez, J., Fisk, A. D., & Rogers, W. A. (2006). Don't let me down: The role of operator expectations on human-automation interaction. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (2345-2349). Santa Monica, CA: Human Factors and Ergonomics Society.

Merriam-Webster Online. (2006). Retrieved September 14, 2006, from http://www.m-w.com/dictionary/trust

Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors, 43*, 563-572.

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors, 46,* 196-204.

Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics, 37,* 1905-1922.

Norman, D. A., & Bobrow, D. G. (1976). On the analysis of performance operating characteristics. *Psychological Review, 83,* 508-510.

Riley, V. (1996). Operator reliance on automation: Theory and data. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 19-35). Mahwah, NJ: Lawrence Erlbaum.

Rogers, W. A., Bertus, E. L., & Gilbert, D. K. (1994). Dual-task assessment of age differences in automatic process development, *Psychology and Aging, 9,* 398-413.

Rosenthal, R. (1966). *Experimenter effects in behavioral research.* New York: Appleton-Century-Crofts.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom.* New York: Holt, Rinehart, and Winston.

Rothbart, M., Evans, M., & Fulero, S. (1979). Recall for confirming events: Memory processes and the maintenance of social stereotypes. *Journal of Experimental Social Psychology, 15,* 343-355.

Salthouse, T. A. (1992). Working-memory mediation of adult age differences in integrative reasoning. *Memory & Cognition, 20,* 413-423.

Sanchez, J., Fisk, A. D., & Rogers, W. A. (2004). Reliability and age-related effects on trust and reliance of a decision support aid. In *Proceedings of the Human Factors Society 48th Annual Meeting* (pp. 586-589). Santa Monica, CA: Human Factors and Ergonomics Society.

Sanchez, J. (2005) *Human-automation interaction: factors that affect human behavior and system performance.* Unpublished manuscript, Georgia Institute of Technology.

Sanchez, J. (2006). *Factors that affect trust and reliance on an automated aid.* Unpublished doctoral dissertation, Georgia Institute of Technology, Atlanta.

Sanchez, J., Ezer, N., Rogers, W. A., & Fisk, A. D. (2006). Estimating reliability of automated aids: Effects of age and system reliability changes. *Cognitive Technology, 11,* 5-13.

Siedlecki, K. L., Salthouse, T. A., & Berish, D. E. (2005). Is there anything special about aging of source memory? *Psychology and Aging, 20,* 19-32.

Snyder, M. L., & Frankel, A. (1976). Observer bias: A stringent test of behavior engulfing the field. *Journal of Personality and Social Psychology, 34,* 857-864.

Shipley, W. C. (1986). *Shipley Institute of Living Scale.* Los Angeles: Western Psychological Services.

Stephan, W. G. (1985). Intergroup relations. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology: Vol. 2,* (3rd ed., Vol. 2, pp. 599-658). United States of America: Newbery Award Records, Inc.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale III. (3rd Ed.).* San Antonio, TX: The Psychological Corporation.

Wickens, C. D., & Carswell, C. M. (2006). Information processing. *In G. Salvendy (Ed.), Handbook of Human Factors and Ergonomics: Third Edition (pp. 1570-1596).* Hoboken, NJ: John Wiley & Sons, Inc.

Wickens, C. D., & Dixon, S. (2002) *Workload demands of remotely piloted vehicle supervision and control: (I) Single vehicle performance* (Tech. Rep. No. AHFD-

02-10/MAD-02-1). Savoy, Illinois: University of Illinois at Urbana-Champaign, Aviation Human Factors Division, Institute of Aviation.

Wickens, C. D., Helleberg, J., & Xu, X. (2002). Pilot maneuver choice and workload in free flight. *Human Factors, 44,* 171-188.

Wickens, C. D., & Xu, X. (2002). *Automation trust, reliability and attention* (Tech. Rep. No. AHFD-02-14/MAAD-02-2). Savoy, Illinois: University of Illinois at Urbana-Champaign, Aviation Human Factors Division, Institute of Aviation.

Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: the effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomic Science, 2,* 352-367.